MULTIVARIATE MODELING FOR A MULTIPLE STAGE, MULTIPLE OBJECTIVE GREEN BUILDING FRAMEWORK

by

PIN KUNG

Presented to the Faculty of the Graduate School of

The University of Texas at Arlington in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT ARLINGTON

August 2012

ACKNOWLEDGEMENTS

When I was a child, my dream was to become an architect. Although the dream did not come true, I am still interested in building design. I would like to thank my research advisor, Dr. Victoria Chen for her guidance and support and for giving me this opportunity to use my industrial engineering background on green building research.

I am grateful to my committee members, Dr. Bill Corley and Dr. Li Zeng, for their suggestions on my dissertation. I would also like to acknowledge green building expert, Anthony Robinson, who gave me valuable research feedback and provided numerous conference paper opportunities.

I would like to thank all my colleagues at the COSMOS Lab, especially Dr. Hui-Chiao Jen, Aera Leboulluec, John Dickson, and Smriti Neogi, for your friendships. Our time together has been fun and I thank you for that. I also would like to thank Siddhi Morje who helped me with my literature search and found the software, ATHENA.

I would like to thank UTA writing center tutors, Abby Allen and Logan Hanson for their help on my dissertation.

I would like to express my gratitude to Karen Collins for taking time to read FriendSpeak books with me on Sundays and for giving me helpful suggestions for my research.

Finally, I would like to thank my parents and my brother for their encouragement. In particular, I deeply thank my dear father who shared his architectural experience with me. Thank you for your endless support.

July 12, 2012

ABSTRACT

MULTIVARIATE MODELING FOR A MULTIPLE STAGE, MULTIPLE OBJECTIVE GREEN BUILDING FRAMEWORK

Pin Kung, PhD

The University of Texas at Arlington, 2012

Supervising Professor: Victoria C. P. Chen

Green building is a sustainable concept to reduce environmental impact. Decision-making for green building is a complex task. A multi-stage green building framework will guide future development of a comprehensive multiple stage, multiple objective (MSMO) decision-making framework. The software eQUEST is utilized in a design and analysis of computer experiments (DACE) approach to study building options that potentially impact energy usage and cost metrics. The DACE approach uses experimental design and statistical analysis to uncover multivariate patterns that will provide guidance for green building decisions. The computer experiments execute the green building software tools ATHENA [11] and eQUEST [13]. The experiment uses a Latin hypercube design to combine a mixed-level orthogonal array for discrete variables with a number-theoretic method for continuous variables. To accommodate the mix of discrete and continuous factor variables, the statistical analysis method fits treed regression (TreeReg) [51], TreeMARS [52], categorical TreeReg (CATreeReg) and categorical TreeMARS (CATreeMARS) models, and uses the method of seemingly unrelated regressions (SUR) [26], [27] to estimate the coefficients for a multiple response linear statistical model.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 Background

Green building has become a popular environmental topic in recent years. It is based on the concept of sustainable development to reduce environmental impact. In the past, buildings were designed by architects only; however, the concept of environmental protection now is promoted to construct “green” buildings. A team which has various experts, such as architects, engineers, contractors, designers, consultants and businessmen, is necessary to complete a large green building project together. A useful environmental analytical tool, life cycle assessment (LCA), is used for the stages of the building cycle which includes material exploitation, manufacturing, construction, operation and maintenance, and demolition (Zhang et al. [1], Retzlaff [2]). This is called a “cradle to grave” design. The fundamental goal of green building is to close the loop on the full life cycle of building construction and operation activities, i.e., the “cradle-to-cradle” concept in Figure 1.1.

There are several main organizations that have develop rules to achieve green concepts for any type of building. The United States Environmental Protection Agency (USEPA) [3] is a government department concerned with the environmental issues and protecting human health. The United States Department of Energy (USDOE) [4] considers energy management, focusing on energy efficiency and renewable energy. The main targets for energy efficiency are building and vehicles, and the primary renewable energy sources are solar and wind. The American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE) [5] considers heating, ventilation, air-conditioning, and refrigeration and publishes a series of standards and guidelines to improve energy use in the building.

Leadership in Energy and Environmental Design (LEED), which was developed by the U.S. Green Building Council (USGBC), provides a building rating system and a set of guidelines that is currently used in practice. The LEED rating system now has several versions, including LEED for New Construction (LEED-NC), LEED for Existing Buildings (LEED-EB), LEED for Commercial Interiors (LEED-CI), LEED for Core and Shell (LEED-CS), LEED for Homes (LEED-H), LEED for Neighborhood Development (LEED-ND), LEED for Schools (LEED-S), LEED for Healthcare (LEED-HC) and LEED for Retail (LEED-R) [6]. For example, LEED-NC is a rating system checklist that includes Sustainable Sites, Water Efficiency, Energy and Atmosphere, Materials and Resources, Indoor Environmental Quality, Innovation in Design and Regional Priority. In the section on water efficiency, a new strategy is considered for greywater and wastewater use. The water reduction in landscaping seeks to improve irrigation efficiency. Each LEED version has a different total point’s requirement. In general, LEED certification has four levels, namely Certified, Silver, Gold and Platinum. These levels give guidance to builders and owners for constructing green buildings.

Energy Star [7], which is supported by USEPA and USDOE, is a rating system that focuses on improving energy efficiency for home, lighting and business. It provides suggestions to guide users how to reduce their energy use. NAHBGreen [8] provides another rating system only for residential buildings and provides new materials and techniques for improving houses every year. It also has a research center to test building materials and improve product quality, and experts build sustainable construction using a national green building standard certification which was issued by the NAHB research center.

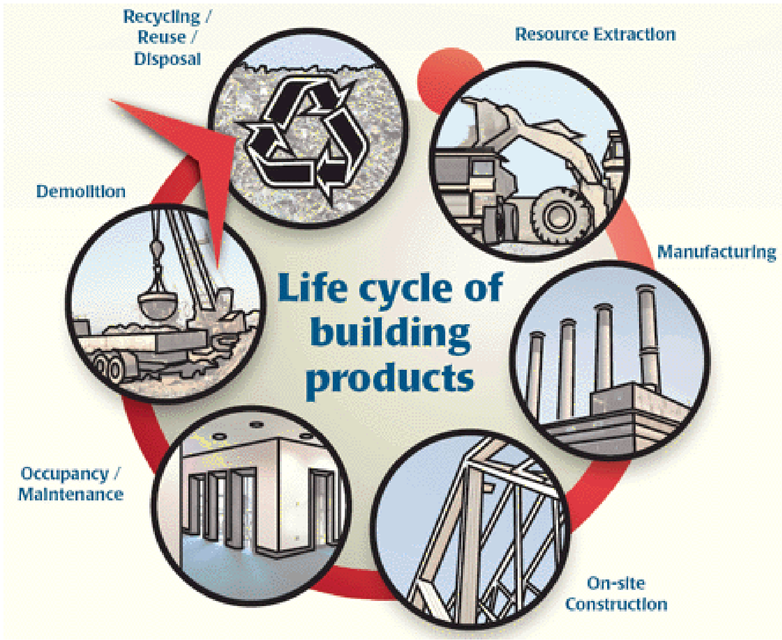


Figure 1.1 Green Building Life Cycle
(Source: http://www.athenasmi.org/wp-content/uploads/
2012/01/LCA_summary_of_four_pages.pdf)

1.2 Research Motivation and Goal

For many complex systems, such as a building, a computer experiment is the only means to comprehensively explore the system. In particular, a building is a complex system that lies within larger complex systems that constitute the development of an urban region, such as the Dallas-Fort Worth metroplex.

Greater understanding is needed on how building options impact green building performance metrics. Operations research focuses on the development of tools for making better, ideally optimal, decisions. This research will present multivariate modeling as part of a multiple stage, multiple objective (MSMO) green building decision-making framework that integrates building expertise with state-of-the-art methods from statistics and operations research to explore, evaluate and select among building technologies. Design and analysis of computer experiments (DACE) [9] is a useful method for studying complex engineering systems, and DACE models will be incorporated into the MSMO framework. The green building design in this research will consider uncertain factors and interactions between the building options affecting multiple performance metrics.

1.3 Organization

There are five chapters in this dissertation. Chapter 1 describes background of green building, motivation and goal. Chapter 2 provides an overview of the previous research, such as green building software tools and existing statistical methods. The proposed MSMO framework and analyses and results are provided in chapter 3 and chapter 4. Finally, conclusions and future work are discussed in chapter 5.

CHAPTER 2

LITERATURE REVIEW

2.1 Green Building Software

Building software [10] is a useful reference tool when constructing a new building. Various software tools can assist in studying components of the framework developed in this dissertation. Three software tools are available to assess performance of building options, specifically ATHENA Impact Estimator for buildings (ATHENA) [11], Building for Environmental and Economic Sustainability (BEES) [12], and the QUick Energy Simulation Tool (eQUEST) [13]. Discussion on software is provided below.

2.1.1 ATHENA

ATHENA is a life cycle tool that assists with making decisions about the selection of material mixes. It is a life cycle tool which provides a cradle-to-grave process for the entire building, where the performance of the building is represented by both life cycle cost (LCC) and environmental impact. Thus, users can use various design options to determine how to decrease environmental impacts and costs. The eight impact measures are fossil fuel consumption, acidification potential, global warming potential, human health respiratory effects potential, ozone depletion potential, smog potential, eutrophication potential, and weighted resource use. Moreover, ATHENA provides inputs for different materials and design options, and it allows users to change designs, use different materials, and make side-by-side comparisons. ATHENA was previously used in the work of Wang et al. [14].

2.1.2 BEES

BEES is based on a life cycle assessment approach for obtaining economic and environmental performance results. It is a free building tool and now provides online web usage for users. Economic and environmental performance are combined into an overall performance measure by using the American Society for Testing and Materials (ASTM) standard. BEES product data contain raw materials, manufacturing, transportation, installation, use, and end of life, and the life cycle cost method covers the costs of initial investment, replacement, operation, maintenance and repair, and disposal. BEES was used in the work by Castro-Lacouture et al. [15].

2.1.3 eQUEST

The software tool eQUEST is a powerful building tool for energy simulation and combines the building energy analysis program DOE-2, graphics and three wizards, namely Schematic Design (SD) Wizard, Design Development (DD) Wizard and Energy Efficiency Measure (EEM) Wizard (Figure 2.1). It was developed together by Lawrence Berkeley National Laboratory and J.J. Hirsch and Associates, under funding from the USDOE and the Electric Power Research Institute. It is reliable and affordable for a broader base of design and buildings professionals. There are 667 of these long-term average weather files, and about 300 North American locations are considered in this software. Heating, Ventilating, and Air Conditioning (HVAC) systems can be modeled using eQUEST. eQUEST is designed to study an entire building and the concept of integrated energy design to construct energy-efficient buildings. In other words, it is designed to provide a whole building analysis to owners, designers or operators, and building designers. Whole building energy modeling consists of a computer program to analyze the annual energy consumption in the buildings. The building construction and operating parameters include building envelope, internal gains, occupancy schedules, and building systems to calculate energy consumption.

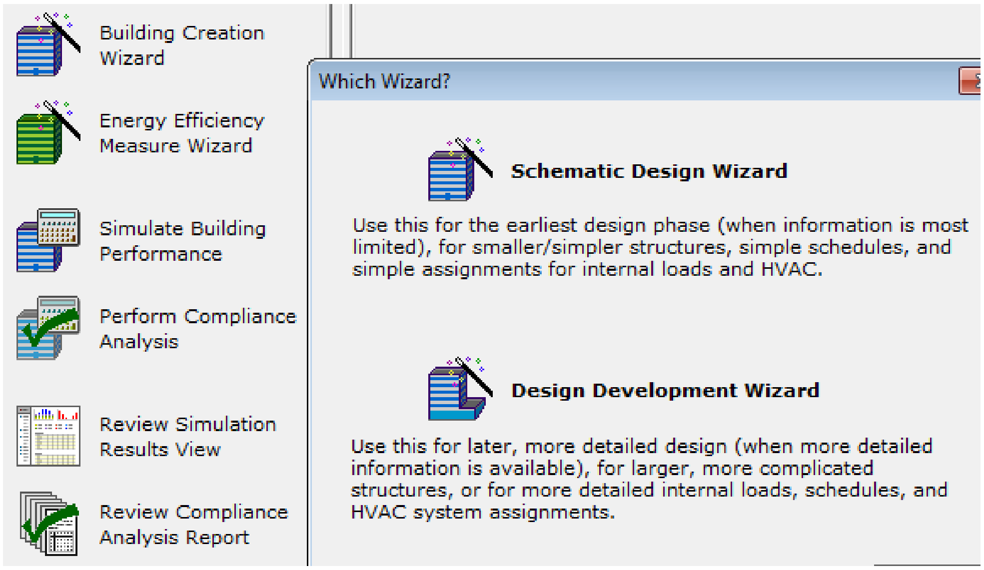


Figure 2.1 eQUEST Software Tool

Figure 2.2 shows the general information. There are 41 screens, including building type, building geometry, construction types, window sizes, door sizes, glass types, activity areas, building operation schedules, HVAC system types, power and efficiencies, water heating types, and so on. Some screens are dependent on previous screen selections. After the 41 screens are completed, the Energy Efficiency Measure Wizard Run should be selected to decide performance metric categories before the simulation. Finally, the reports of annual building summary have three parts. Various energy types, peak, utility cost, and life cycle cost (LCC) include total annual results in the upper part, incremental annual savings in the middle part, and cumulative annual savings in the bottom part. eQUEST provides many detailed screens for builders to input a variety of options, and Design Development Wizard provides more flexible options to handle complex buildings.

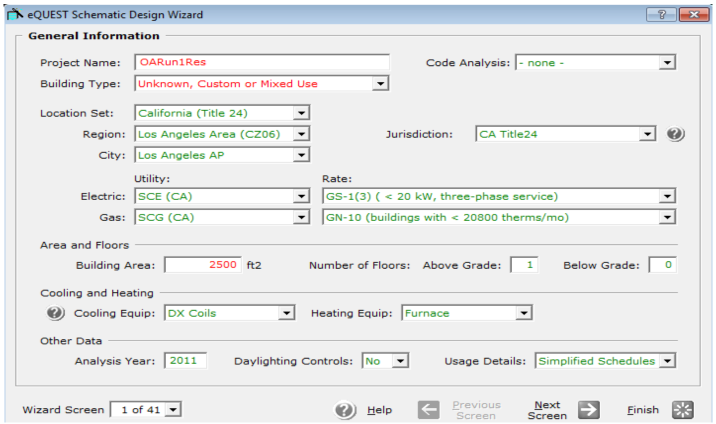


Figure 2.2 eQUEST General Information

2.1.4 Comparison

Among the three software tools, BEES and ATHENA do not enable study of building orientation, which is generally acknowledged to be an important aspect of green building. BEES is not as flexible or comprehensive as the other two; however, eQUEST focuses on energy use and does not consider environmental impacts like BEES and ATHENA. In Figure 2.3, ATHENA and BEES are included within the sustainability circle, where the performance criteria are life cycle environmental impact and life cycle cost, and eQUEST is included within the energy efficiency circle, where the performance criteria are energy use and life cycle cost. To study the intersection of these criteria is the main goal. For the current study, only eQUEST is used in the dissertation.

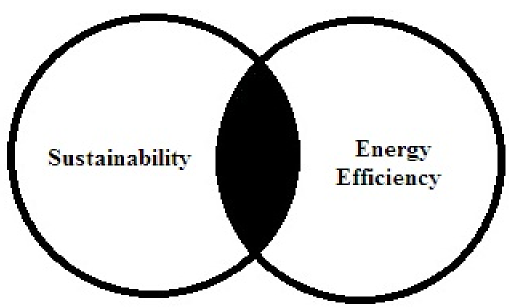


Figure 2.3 Green Building Evaluation for Three Software Tools

2.2 Green Building Optimization

The goal of green building is achieving greater energy efficiency and minimizing environmental impacts. Several researchers have shown that optimization can be applied in the building research. Using LEED goals, Castro-Lacouture et al. [15] used mixed integer optimization to select appropriate materials, so as to maximize the points achievable for a LEED-based rating system under a given budget. Wang et al. [14] used a multiple objective genetic algorithms optimization to study building orientation and aspect ratio, windows, walls, and roofs, so as to balance life cycle cost and environmental impact. A genetic algorithms optimization appears to provide the ultimate flexibility needed for green building decision-making, but a genetic algorithm is not guaranteed to yield globally optimal solutions in practice, and the algorithm can be very computationally expensive. Nielsen [16] developed a software tool to design an entire building, so as to minimize life cycle cost subject to constraints related to energy usage and indoor environment. Limitations of the tool include the use of simple mathematical models to assess performance objectives and the use of existing optimization software. Osman et al. [17] conducted a life cycle analysis using a linear program to separately optimize three performance metrics: cost, life cycle global warming potential, and tropospheric ozone precursor potential. They also varied the parameters of the linear program to empirically draw Pareto frontiers illustrating the trade-offs between the performance metrics. Hasan et al. [18] combined the software GenOpt, which minimizes a cost function that is evaluated by an external simulation, and the IDA Indoor Climate and Energy program, which calculates heating energy consumption, to seek the minimum life cycle cost. BEopt [19] is the other software which considers the optimal building design to achieve the goal of zero net energy. The Center for Sustainable Systems [20] had similar projects, such as life cycle design and life cycle optimization. All above existing methods did not consider uncertainty or dependencies between the building decision options and a mix of discrete and continuous decision variables for MSMO decisions, and this dissertation will provide useful methods and discuss the results in chapter 3 and chapter 4. The further building optimization using MSMO will be described in chapter 5, but optimization will not be used for the current study.

2.3 Statistical Analysis-Existing Methods

Statistics is a useful analysis tool which provides various methods to solve different questions. At first, researchers studied statistical models with single response. Later, multivariate analyses handling multiple responses were applied extensively in statistical research [21]. Some statistical methods, both single response and multiple responses, are described as follows.

2.3.1 Multiple Linear Regression

2.3.1.1 Single Response

Multiple linear regression (MLR) [21] is the most widely used method in engineering statistics. It is to discuss the model relationship between predictor variables and a dependent variable, and "linear" denotes linear in the unknown model parameters (or coefficients). The MLR model is

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{2.1}$$

where $\mathbf{Y}$ is a vector of the response variable observations, $\mathbf{X}$ is a matrix of the values of the predictor variables, $\boldsymbol{\beta}$ is a vector of unknown model parameters, and $\boldsymbol{\varepsilon}$ is a vector of random error terms. Thus, the model can be written as

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & \cdots & X_{1,p-1} \\ 1 & X_{21} & \cdots & X_{2,p-1} \\ \vdots & \vdots & & \vdots \\ 1 & X_{n1} & \cdots & X_{n,p-1} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}. \tag{2.2}$$

There are n observations and $p-1$ predictor variables in the MLR model. Using ordinary least squares (OLS), the estimator for $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{Y}, \tag{2.3}$$

and error terms are assumed

$$E(\boldsymbol{\varepsilon}) = \mathbf{0} \text{ and } Cov(\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{I}. \tag{2.4}$$

A full analysis, include parameter tests, confidence intervals, analysis of variance, prediction, and residual analysis, can be found in [21].

Stepwise regression [22] which is an approach for the model selection of predictor variables is the most commonly used in single response. Since stepwise variable selection produces biased regression coefficients, Copas [23] described a shrinkage method which is to shrink estimated coefficients towards 0, and this method can reduce the prediction mean squared error. Tibshirani [24] developed a new method, least absolute shrinkage and selection operator (lasso), and indicated the coefficients need to be shrunk more to improve the performance of subset selection and ridge regression.

2.3.1.2 Multiple Responses

2.3.1.2.1 Seemingly Unrelated Regressions

Zellner [25] developed a new technique, Seemingly Unrelated Regressions (SUR). The key point in this technique is that using a method that is designed for multiple responses, particularly when the responses are correlated is more efficient than using OLS separately on the different responses when predictions have higher standard errors. If the responses are not correlated, it is appropriate to use OLS separately. Let the matrix $\mathbf{Y}$ consist of M response variables each with $n \times 1$ vectors of observations, let $\mathbf{X}$ be the block-diagonal matrix for the predictors as in (2.1), let $\boldsymbol{\beta}$ be a matrix with M sets of unknown parameters, and let $\boldsymbol{\varepsilon}$ be a matrix of M error terms each with $n \times 1$ vectors. Thus, SUR model can be written as

$$\begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \\ \vdots \\ \mathbf{Y}_M \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{X}_M \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \vdots \\ \boldsymbol{\beta}_M \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \vdots \\ \boldsymbol{\varepsilon}_M \end{bmatrix}. \tag{2.5}$$

The variance-covariance matrix of error terms is

$$\begin{bmatrix} \sigma_{11}\mathrm{I}_n & \sigma_{12}\mathrm{I}_n \cdots \sigma_{1M}\mathrm{I}_n \\ \sigma_{21}\mathrm{I}_n & \sigma_{22}\mathrm{I}_n \cdots \sigma_{2M}\mathrm{I}_n \\ \vdots & \vdots \qquad \vdots \\ \sigma_{M1}\mathrm{I}_n & \sigma_{M1}\mathrm{I}_n \cdots \sigma_{MM}\mathrm{I}_n \end{bmatrix} = \mathbf{\Sigma} \otimes \mathbf{I}_n , \tag{2.6}$$

where $\mathbf{\Sigma}$ is a $M \times M$ matrix of σ. It is desired that the error standard deviation estimator $\hat{\sigma}$ be unbiased, so Henningsen and Hamann [26] discussed methods to modify the degrees of freedom associated with the errors when each equation has the different number of predictors. They provided two approached to compute the residual covariance matrix for response variables $\mathbf{Y}_i$ and $\mathbf{Y}_j$. The first approach is frequently used in the software and the equation is

$$\hat{\sigma}_{ij} = \frac{\mathbf{Y}_i^{\mathrm{T}}[\mathbf{I}_{\mathrm{n}} - \mathbf{X}_i(\mathbf{X}_i^{\mathrm{T}}\mathbf{X}_i)^{-1}\mathbf{X}_i^{\mathrm{T}}][\mathbf{I}_{\mathrm{n}} - \mathbf{X}_j(\mathbf{X}_j^{\mathrm{T}}\mathbf{X}_j)^{-1}\mathbf{X}_j^{\mathrm{T}}]\mathbf{Y}_j}{\sqrt{(n-k_i)(n-k_j)}} , \tag{2.7}$$

where $i, j = 1, 2, \ldots, M$, k_i and k_j are the numbers of predictor variables, the residuals are computed by OLS, and $\otimes$ denotes Kronecker product. The SUR model vector of parameter estimates is

$$\hat{\boldsymbol{\beta}}_{\mathrm{SUR}} = [\mathbf{X}^{\mathrm{T}}(\hat{\mathbf{\Sigma}}^{-1} \otimes \mathbf{I}_n)\mathbf{X}]^{-1}\mathbf{X}^{\mathrm{T}}(\hat{\mathbf{\Sigma}}^{-1} \otimes \mathbf{I}_n)\mathbf{Y} . \tag{2.8}$$

The SUR's error terms are correlated across equation, and it is more efficient than OLS when the different response variables are correlated.

Based on the idea of Zellner [25], Shah et al. [27] considered an extension of the SUR technique in multiple-response response surface methodology (RSM) problems. In the paper, Shah et al. [27] showed SUR and OLS will have the same parameter estimates when the models' forms are the same and when the error terms for the different response variable models are uncorrelated. In a numerical example, they showed the parameter estimates of SUR were more accurate than OLS when responses were correlated.

2.3.1.2.2 Curds and Whey Procedure

Breiman and Friedman [28] developed a curds and whey (C&W) method which is a shrinking method and can be used for multi-response problems to reduce prediction errors when there are correlations between responses. They first supposed the error terms are independent and identically distributed $N(0,\sigma^2)$. The response variables are correlated because they use the same predictor variables, so a more accurate predictor $\tilde{Y}$ is considered in place of $\hat{Y}$. The multi-response model is written as

$$\tilde{Y}_i = \overline{Y}_i + \sum_{k=1}^{q} \beta_{ik} (\hat{Y}_k - \overline{Y}_k); i = 1, 2, \ldots, q, \tag{2.9}$$

where

$$\hat{Y}_i = \overline{Y}_i + \sum_{j=1}^{p} \hat{\alpha}_{ij} (x_j - \overline{x}_j), \tag{2.10}$$

$$\hat{\alpha}_{ij} = \arg\min\{\sum_{l=1}^{n} [Y_{il} - \overline{Y}_i - \sum_{j=1}^{p} \alpha_j (x_{jl} - \overline{x}_j)]^2\}. \tag{2.11}$$

Thus, the model can be rewritten as

$$\begin{aligned} \tilde{Y}_1 &= \overline{Y}_1 + \sum_{k=1}^{q} \beta_{1k} \sum_{j=1}^{p} \hat{\alpha}_{1j} (x_j - \overline{x}_j) \\ \tilde{Y}_2 &= \overline{Y}_2 + \sum_{k=1}^{q} \beta_{2k} \sum_{j=1}^{p} \hat{\alpha}_{2j} (x_j - \overline{x}_j). \\ &\vdots \\ \tilde{Y}_q &= \overline{Y}_q + \sum_{k=1}^{q} \beta_{qk} \sum_{j=1}^{p} \hat{\alpha}_{qj} (x_j - \overline{x}_j) \end{aligned} \tag{2.12}$$

If the predictor and response variables are considered to be centered, then equation (2.9) can be written as

$$\tilde{\mathbf{Y}} = \mathbf{B}\hat{\mathbf{Y}}. \tag{2.13}$$

The optimal shrinkage matrix $\mathbf{B}^*$ is obtained from the estimator $\mathbf{B}$, and $\mathbf{B}^*$ is a diagonal matrix that can be expressed as

$$\mathbf{B}^* = \mathbf{T}^{-1}\mathbf{D}\mathbf{T}, \tag{2.14}$$

where

$$\mathbf{D} = \mathrm{diag}\{d_1, \ldots, d_q\} \tag{2.15}$$

is composed of the shrinkage factors d_i using the squared canonical correlations c_i^2 as follows

$$d_i = \frac{c_i^2}{c_i^2 + r(1-c_i^2)}; i = 1, 2, \ldots, q, \tag{2.16}$$

where p predictor variables and n observations are

$$r = \frac{p}{n}, \tag{2.17}$$

so the values of d_i are between 0 and 1.

In Srivastava and Solanky [29], they described the C&W estimator in detail. The OLS predictor in C&W is based on equation (2.1), and the formula of the least squares estimator $\hat{\mathbf{A}}$ is the same as equation (2.3); however, $\hat{\mathbf{A}}$ now is a $p \times q$ matrix of unknown parameters. Therefore, the C&W estimator is

$$\hat{\boldsymbol{\beta}}_{CW} = \hat{\mathbf{B}}_{CW}\hat{\mathbf{A}}^{\mathrm{T}}, \tag{2.18}$$

where the shrinkage estimator is

$$\hat{\mathbf{B}}_{CW} = [\mathbf{I}_q + r\mathbf{S}(\mathbf{Y}^{\mathrm{T}}\mathbf{M}\mathbf{Y})]^{-1}, \tag{2.19}$$

and

$$\mathbf{S} = \mathbf{Y}^{\mathrm{T}}(\mathbf{I}_n - \mathbf{M})\mathbf{Y}, \tag{2.20}$$

and

$$\mathbf{M} = \mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}. \tag{2.21}$$

Thus, $\mathbf{B}^*$ can be estimated by $\hat{\mathbf{B}}_{CW}$ which is a $q \times q$ matrix. Moreover, Xu et al. [30] provided a new method which combined curds and whey with partial least squares (PLS), called PLS-C&W. They compared the performance metrics of PLS models with their corresponding C&W models, and the results showed the PLS-C&W is the better method.

2.3.2 Regression Trees

2.3.2.1 Single Response

Breiman et al. [31] developed a classification and regression tree (CART) method which fits piecewise constant models at each node of the tree. In the software R [32], there are two packages, ”tree” [33] for classification and regression trees and “rpart” [34] for recursive partitioning and regression trees. In Friedman [35], CART has forward and backward stepwise algorithms. The CART forward stepwise algorithm uses recursive partitioning (Algorithm 1). Another tree algorithm was developed by Loh [36], using a generalized, unbiased interaction detection and estimation (GUIDE) algorithm which is based on a chi-square test for a piecewise-constant fit and splitting at each node of the tree, or piecewise linear models which use a chi-square test for linear fit and splitting at each node of the tree. Categorical and numerical predictor variables are also allowed in GUIDE. Kim et al. [37] developed an algorithm which is an extension to GUIDE in which they used stepwise regression to fit a linear model. Compared with all GUIDE algorithms in this paper, GUIDE with piecewise stepwise linear model had small geometric means of root-mean-square error (RMSE).

2.3.2.2 Multiple Responses

De’ath [38] developed a multivariate regression trees (MRT) technique which is an extension of univariate regression trees (URT) from Breiman et al. [31]. Although the growing and pruning in MRT are more complicated than URT, most the methods of MRT are the same as URT. For example, MRT and URT use cross-validation to select tree size and use the same method to split at the nodes of the tree. The R package mvpart [39] which is an extension of rpart is provided by De’ath. The average response values are shown at each terminal node.

GUIDE is the other software tool which can be used in multiple response regression tree models, and this algorithm fits piecewise constant models for multi-response data (Loh [40]). In the manual, a section is provided for user to do step by step and understand how GUIDE works on multiple response data. In general, mvpart is more comprehensive than GUIDE.

2.3.3 Multivariate Adaptive Regression Splines

2.3.3.1 Single Response

Friedman [35] in 1991 proposed multivariate adaptive regression splines (MARS), which uses a forward stepwise procedure (Algorithm 2) and a backward stepwise procedure (Algorithm 3) to fit a flexible model to predict a single response variables as a function of multiple predictor variables (or covariates). The MARS linear approximation can be expressed as

$$\hat{f}(\mathbf{x}) = \beta_0 + \sum_{m=1}^{M} \beta_m B_m(\mathbf{x}), \tag{2.22}$$

where the vector of predictor variables is $\mathbf{x} = (x_1, x_2, \cdots, x_p)^{\mathrm{T}}$, M is the number of basis functions, β_m is the unknown coefficient for the m th basis function, and B_m is the m th basis function. A basis function is a product of truncated linear functions, where a single truncated linear function is defined as:

$$(x-t)_+ = \begin{cases} x-t & x > t \\ 0 & otherwise \end{cases} \tag{2.23}$$

or

$$(t-x)_+ = \begin{cases} t-x & x < t \\ 0 & otherwise \end{cases}. \tag{2.24}$$

The pair of functions $(x-t)_+$, $(t-x)_+$ is called reflected pair, and t is called a knot of the basis functions. A basis function can then be written as:

$$B_m(x) = \prod_{k=1}^{K_m} [s_{km}(x_{v(k,m)} - t_{km})]_+ \,, \tag{2.25}$$

where K_m is the number of truncated linear functions multiplied in the m th basis function, $x_{v(k,m)}$ is the input variable corresponding to the k th truncated linear function in the m th basis function, s_{km} can be +1 or −1, and t_{km} is the knot value corresponding to $x_{v(k,m)}$. A main effect basis function involves only one truncated linear function, while an interaction involves products.

The generalized cross-validation criterion is used to quantify how well a MARS model fits the data:

$$LOF(\hat{f}_M) = GCV(M) = \frac{\frac{1}{N}\sum_{i=1}^{N}[y_i - \hat{f}_M(x_i)]^2}{\left[1 - \frac{C(M)}{N}\right]^2} \,, \tag{2.26}$$

where the complexity cost function is

$$g(M) = Number\ of\ Terms + Penalty \times \frac{Number\ of\ Terms - 1}{2} \,, \tag{2.27}$$

and the penalty on complexity is usually 2 or 3 (Friedman [35] and MARS from Wikipedia [41]). Based on the MARS algorithm from Friedman [35], Sekulic and Kowalski [42] provided four different examples, namely additive model, interaction model, complex model and exponential transformation model to show their MARS performance.

2.3.3.2 Multiple Responses

Hastie et al. [43] used the method of a flexible discriminant analysis (FDA) in multi-response MARS. FDA, which can be used in nonlinear classification, is based on the concept of Fisher's linear discriminant analysis (LDA) [44]. The multi-response MARS linear approximations can be expressed as

$$\begin{aligned}
\hat{Y}_1 &= \hat{f}_1(\mathbf{x}) = \beta_{01} + \sum_{m=1}^{M} \beta_{m1} B_m(\mathbf{x}), \\
\hat{Y}_2 &= \hat{f}_2(\mathbf{x}) = \beta_{02} + \sum_{m=1}^{M} \beta_{m2} B_m(\mathbf{x}), \\
&\vdots \\
\hat{Y}_q &= \hat{f}_q(\mathbf{x}) = \beta_{0q} + \sum_{m=1}^{M} \beta_{mq} B_m(\mathbf{x}),
\end{aligned} \tag{2.28}$$

where there are q response variables. In this equation, the responses have the same basis functions $B_m(\mathbf{x})$ which are considered to be fixed functions in the FDA procedure after they have been chosen by a MARS procedure, but they use the different coefficients ($\beta_{m1}, \beta_{m2}, \ldots, \beta_{mq}$) which can be obtained by ordinary least squares, penalized least squares, or other methods. Therefore, q simultaneous models are estimated.

These models are built and pruned the same way as the single response MARS. The only difference is that the residual sum of squares (RSS) and generalized cross validation (GCV) criterion involve sums across all q response variables. In R, the package "earth" conducts multi-response MARS, but does not demonstrate as good results as building the models independently since the "earth" multi-response version forces the same set of basis functions for all q models.

2.3.4 Projection Pursuit Regression

2.3.4.1 Single Response

Projection pursuit, which was develop by Friedman and Tukey [45], seeks to find the most useful projections in the multidimensional predictor space and utilize the projection direction to pursuit the maximum projection index. They describe a one-dimensional projection index $I(k)$ that has a projection axis k on a one-dimensional line, and a two-dimensional projection index $I(k,l)$ that has two projection axes k and l on a two-dimensional plane.

Three-dimensional and higher dimensional projection pursuit (PP) are discussed in Friedman [46]. The basic linear model estimate is

$$\hat{Y} = f(\mathbf{x}) = \alpha_0 + \sum_{j=1}^{p} \alpha_j x_j \ , \tag{2.29}$$

where the vector of predictor variables is $\mathbf{x} = (x_1, x_2, \cdots, x_p)$, α_0 is intercept term and α_j is slope term. f is a smooth function based on local averaging and is fitted based on the least squares criterion

$$L_2(\alpha_0 \cdots \alpha_p) = E[Y - \hat{Y}]^2 \, . \tag{2.30}$$

Friedman and Stuetzle [47] used an extension which considered PP for regression problems, i.e. PPR. PPR only considers a single response and extends the basic linear model and the approximation equation to

$$\hat{Y} = \sum_{m=1}^{M} f_m (\sum_{j=1}^{p} \alpha_{jm} x_j) \ , \tag{2.31}$$

where $f_m(\sum_{j=1}^{p} \alpha_{jm} x_j)$ is called a ridge function, α_{mj} is a projection direction, x is a loading vector, and f_m is a smooth transfer function. Friedman [48] described the advantages and disadvantages of the PPR algorithm. The least squares criterion for PPR is

$$L_2(\alpha_1^T \cdots \alpha_M^T, f_1 \cdots f_M) = E[Y - \hat{Y}]^2 \, . \tag{2.32}$$

2.3.4.2 Multiple Responses

SMART is a generalization of PPR to handle q response variables ($Y_i, 1 \le i \le q$) and p predictor variables ($x_j, 1 \le j \le p$). The approximation equations can be expressed as

$$
\begin{aligned}
\hat{Y}_1 &= \overline{Y}_1 + \sum_{m=1}^{M} \beta_{1m} f_m (\sum_{j=1}^{p} \alpha_{jm} x_j), \\
\hat{Y}_2 &= \overline{Y}_2 + \sum_{m=1}^{M} \beta_{2m} f_m (\sum_{j=1}^{p} \alpha_{jm} x_j), \\
&\vdots \\
\hat{Y}_q &= \overline{Y}_q + \sum_{m=1}^{M} \beta_{qm} f_m (\sum_{j=1}^{p} \alpha_{jm} x_j),
\end{aligned}
\tag{2.33}
$$

where $\overline{Y}_i = E(Y_i)$, $E(f_m) = 0$, $E(f_m^2) = 1$, $\sum_{j=1}^{p} \alpha_{jm}^2 = 1$, and β_{im}, α_{jm} and the identical f_m models are estimated by least squares. In SMART, the q response variables will share a set of functions, so the basis functions $B_m(\mathbf{x})$ equals $f_m(\sum_{j=1}^{p} \alpha_{jm} x_j)$ (Hastie et al. [43]). In general, when $q = 1$ in SMART, SMART and PPR have the same form. However, SMART and PPR use a different way to choose their estimates. If the predictor variables have the high associations, SMART and PPR will be different models (Friedman [49] and Frank [50]). The least squares criterion for SMART is

$$
L_2(\alpha_1^T \cdots \alpha_M^T, \beta_1^T \cdots \beta_M^T, f_1 \cdots f_M) = \sum_{i=1}^{q} W_i E[Y_i - \hat{Y}_i]^2, \tag{2.34}
$$

where W_i is the response weight that provides the balance for the importance of the different responses can be specified by the user. Friedman [49] also provided W_i can equal to 1 divided by variance $\mathrm{var}(Y_i)$ or can rescale the response variables Y_i to have the same variance if each response has the same importance in (2.34).

2.3.5 Hybrid Tree Method

2.3.5.1 Single Response

2.3.5.1.1 Treed Regression

The CART model specifies that the regression functions are constants at the terminal nodes, and treed regression model [51] considers linear regressions at the terminal nodes instead. The tree portion can handle categorical and numerical variables, while the linear regression models are applied only to the numerical variables. In the single response case, Loh [40] developed a GUIDE algorithm and used piecewise linear models which are identical in concept to that of treed regression.

2.3.5.1.2 Treed Multivariate Adaptive Regression Splines

Sahu [52] followed the same concept as treed regression, but used MARS at the terminal nodes instead of regression. He called his model TreeMARS. Similar to treed regression, the tree portion can handle categorical and numerical variables, but the MARS portion only handles numerical variables well (there is an option for binary variables in MARS, but it does not model well). TreeMARS can be implemented by using only the categorical variables in the tree or by using both categorical and continuous variables in the tree. Results from Sahu indicate that using only the categorical variables in the tree is preferable since MARS is superior at modeling the numerical variables.

2.3.6 Mahalanobis-Taguchi

2.3.6.1 Multiple Responses

Pan et al. [53] described a Mahalanobis-Taguchi system (MTS) [54] that combined the method of the scaled Mahalanobis distance in a multidimensional system with the Taguchi method. This system can be used to optimize the system and predict performance. Since the traditional multidimensional system uses multi-response variables, which are mutually independent, mistakes will happen in determining the important factors for response variables. Thus, they considered the concept of Mahalanobis-Taguchi-Gram Schmidt (MTGS), which uses the Gram-Schmidt orthonormalization process in MTS to compute the scaled Mahalonobis distances. They also provided four steps: (1) perform the standardization, (2) perform the orthogonalization, (3) compute the Mahalonobis distance, and (4) perform analysis of variance (ANOVA) and regression analysis. When they calculated the Mahalonobis distance, the equation

$$Y_i = \sqrt{MD_i} \tag{2.35}$$

can be used to convert the multiple response variables into a single response variable (Y), where MD denotes Mahalonobis distance, and $i = 1, 2, \ldots, n$ denotes the ith run. Therefore, the method of MD can be used to convert multiple responses into a single response.

2.3.7 Partial Least Squares or Projection to Latent Structures

2.3.7.1 Single Response

PLS is a correlation-based technique that was originally developed by Hermann Wold. Manne [55] described the PLS1 algorithm for predicting a single response variable in chemometrics. The PLS2 algorithm allows multiple response variables and is widely used in research.

2.3.7.2 Multiple Responses

In the PLS2 model, the outer model decomposes the matrix of X variables and the matrix of Y variables into the form

$$\mathbf{X} = \mathbf{t}\mathbf{P}^{\mathrm{T}} + \mathbf{E}, \tag{2.36}$$

$$\mathbf{Y} = \mathbf{u}\mathbf{Q}^{\mathrm{T}} + \mathbf{F}, \tag{2.37}$$

where $\mathbf{X}$ is an $N \times K$ matrix, $\mathbf{Y}$ is an $N \times M$ matrix, $\mathbf{P}$ is a input loading vector, $\mathbf{Q}$ is a output loading vector, $\mathbf{E}$ and $\mathbf{F}$ are residual matrices, and $\mathbf{t}$ is a input scores vector

$$\mathbf{t} = \mathbf{X}\mathbf{w}, \tag{2.38}$$

where $\mathbf{w}$ denotes weight, and $\mathbf{u}$ is output scores vectors. The inner relation contains a linear inner model and a nonlinear inner model. The nonlinear inner model between $\mathbf{t}$ and $\mathbf{u}$ is

$$\mathbf{u} = f(\mathbf{t}) + \mathbf{e} . \tag{2.39}$$

Wold et al. [56] developed the original quadratic PLS. The nonlinear PLS model with quadratic inner relation is

$$\hat{\mathbf{u}} = \hat{f}(\mathbf{t}) = c_0 + c_1\mathbf{t} + c_2\mathbf{t}^2 , \tag{2.40}$$

where the coefficients c_0, c_1 and c_2 are estimated by least squares.

Baffi et al. [57] modified the nonlinear quadratic PLS algorithm to handle non-linear data. They used three weight updating procedures, namely PLS_A, PLS_B and PLS_C (an error based quadratic PLS algorithm). PLS_C was compared with Wold et al. [56] and traditional linear PLS. Qin and McAvoy [58] developed a nonlinear PLS with a neural network function. Baffi et al. [59] proposed error-based neural network PLS algorithms. They developed neural network PLS algorithms with a sigmoid neural network and a radial basis function (RBF) network. For example, the nonlinear PLS with a sigmoid neural network calculates the nonlinear prediction of $\mathbf{u}$ as:

$$\hat{\mathbf{u}} = \omega_2 \cdot \sigma(\omega_1 \cdot \mathbf{t} + \beta_1) + \beta_2 , \tag{2.41}$$

where the weights are ω_1 and ω_2, the biases are β_1 and β_2, and the centered sigmoidal activation function is σ. For a multiple linear regression problem, Wold et al. [60] discussed PLS-regression (PLSR) as a generalization of multiple linear regression that can analyze data with many noisy, collinear predictor variables, and simultaneously model multiple response variables.

CHAPTER 3

MSMO FRAMEWORK FOR GREEN BUILDING

3.1 MSMO Decision-Making Framework

The ultimate research goal is to develop a comprehensive MSMO green building decision-making framework. The multiple objectives correspond to the various performance metrics that are needed to assess sustainability. Based on the information from [61] and in consultation with Mr. Anthony Robinson for this research, the building options were organized into twelve main building categories, shown in Table 3.1: (1) siting options, (2) electrical system, (3) wells and septic system, (4) foundation system, (5) plumbing system, (6) wall system, (7) window system, (8) door system, (9) roof system, (10) ventilation system, (11) heating and cooling system, and (12) landscaping system. These categories constitute the stages of the framework, and each stage has unique technology options. Table 3.1 provides an example listing of options within each stage, where those options available in ATHENA, BEES, and eQUEST are noted in parentheses.

For siting options (stage 1), electrical (stage 2), foundation (stage 4), wall (stage 6), window (stage 7), door (stage 8), roof (stage 9), ventilation (stage 10) and heating and cooling (stage 11) systems, most of their options can be modeled using three software tools. For example, eQUEST has orientation and footprint options in stage 1. eQUEST has AC system option in stage 2, but solar system and both AC and solar system can not be modeled using three software tools. The foundation system includes foundation and floor, and only the concrete ground floor option can be modeled using eQUEST. For stage 3, stage 5 and stage 12, the example options can not be modeled using three software tools. For instance, concrete septic tank and fiberglass septic tank options are listed in stage 3, which are not options in three software tools. For the electrical system (stage 2) and the heating and cooling system (stage 11), options are based on eQUEST defaults.

In comprehensive MSMO framework, the multiple objectives will involve multiple performance metrics quantified by the various software tools, such as life cycle cost, environmental impact, utility cost, and total source energy. These objectives also correspond to the multiple response variables in multivariate statistical analysis. Multivariate modeling from a DACE process will be used to represent relationships between green building options (decision variables) and performance metrics (objectives). Representation of uncertainty in these statistical models will be used to simulate uncertainty within the decision-making framework, as in stochastic optimization [62-63]. The MSMO framework will also handle possible dependencies between options in different stages [64]. For example:

- The choice of a fan system in stage 11 depends on the electrical system selected in stage 2.
- The choice of a sprinkler system in stage 12 may depend on plumbing decisions in stage 5.

Both traditional and green building options are considered in the framework, so as to enable the study of the benefit of green building.

Table 3.1 Stages and Decision Variables for Green Building

Stage	Building Stage with Options
1	**Siting Options** ● Orientation and Footprint (eQUEST)
2	**Electrical System** ● AC System (eQUEST) ● Both AC and Solar System ● Solar System
3	**Wells and Septic System** ● Concrete Septic Tank ● Fiberglass Septic Tank
4	**Foundation System** ● Concrete Ground Floor (eQUEST) ● Concrete Slab on Grade (ATHENA) ● Generic Portland Cement (BEES) ● Steel Foundation System

Table 3.1 – *Continued*

5	**Plumbing System** ● Freshwater System ● Greywater System ● Rainwater Catchment System
6	**Wall System** ● Concrete Wall (ATHENA, BEES, eQUEST) ● Curtain Wall (ATHENA) ● Drywall ● Metal Frame (eQUEST) ● Straw Bale Walls ● Wood Frame (eQUEST)
7	**Window System** ● Clear/Tint Windows (eQUEST) ● Glazed Windows ● Low-e Windows (eQUEST) ● Reflective Windows (eQUEST) ● Wood Frame Windows (ATHENA, eQUEST)
8	**Door System** ● Steel Door (ATHENA, eQUEST) ● Wood Door (eQUEST)
9	**Roof System** ● Concrete Tile Roof (ATHENA, eQUEST) ● Generic Fiber Cement Roof (BEES) ● Roof Surface Materials (eQUEST)
10	**Ventilation System** ● Balanced Ventilation System ● Exhaust Ventilation System ● Supply Ventilation System ● Ventilation-Activity Areas (eQUEST)
11	**Heating and Cooling System** ● Fan System (eQUEST) ● HVAC System (eQUEST)
12	**Landscaping System** ● Sprinkler System

3.1.1 DACE Exploratory Methodology

3.1.1.1 Design of Experiments

In this research, a case study is a single-story residential building. Building options were considered for a 2500 square-foot, one-story, single-family residential low-rise building. Table 3.3 specifies the assumed values for the areas of the residential low-rise building. After carefully reviewing the allowable building options in eQUEST, 46 decision variables are identified in eQUEST (Table 3.2). These 46 variables correspond to factor variables in the statistical analysis. For an exploratory analysis, each factor variable is limited to two settings, specified in

Table 3.4. For example, two footprint dimensions, specifically 100×25 and 50×50, are considered. The settings for the other factor variables were selected from eQUEST input options.

In addition, 68 uncontrollable variables are identified in eQUEST. For example, Occupied Loads-Lighting-2 specifies the lighting load when activity area 2 (bedroom) is occupied. These variables are likely uncontrollable, so they are represented by random variables. Table 3.5 specifies the ranges for these variables, based on eQUEST default values. A uniform distribution over these ranges was used to sample the values of these random variables for the software runs. Note that the indices 1-8 on some variables in Tables 3.3, 3.4 and 3.5, e.g., "Design Max Occupant Density-4," correspond to the eight area types specified in Table 3.3. The unit of design max occupant density is square feet per person, and the unit of design ventilation is cubic feet per minute (CFM) per person.

Table 3.2 46 Decision Variables

Stage	Building Category	46 Factors
1	Siting Options	• Footprint X&Y • Orientation
2	Electrical System	**Based on Default from eQUEST**
3	Wells and Septic System	**No Option in eQUEST**
4	Foundation System	• Ground Floor Construction • Ground Floor Interior Insulation • Ground Floor Cap • Ground Floor Exterior/Cavity Insulation
5	Plumbing System	**No Option in eQUEST**
6	Wall System	• Wall Construction • Exterior Wall Finishes • Exterior Wall Color • Exterior Wall Insulation • Additional Wall Insulation • Interior Wall Insulation
7	Window System	• Window-Glass Category • Windows-Glass Type • Windows-Frame Width • Window Height • Windows-Sill Height • Distance from Window-Overhangs
8	Door System	• Doors Construction • Door Glass Type

Table 3.2 – *Continued*

		• Door Dimension-Height&Width • Door-Frame Width
9	Roof System	• Roof Construction • Exterior Roof Finish • Exterior Roof Color • Exterior Roof Insulation • Additional Roof Insulation • Ceiling Interior Finishes • Ceiling Batt Insulation • Pitched Roof
10	Ventilation System	• Design Max Occupant Density-Residential (General Living Space) • Design Ventilation-Residential (General Living Space) • Design Max Occupant Density-Residential (Bedroom) • Design Ventilation- Residential (Bedroom) • Design Max Occupant Density-Residential (Garage) • Design Ventilation-Residential (Garage) • Design Max Occupant Density-Dining Area • Design Ventilation-Dining Area • Design Max Occupant Density-Kitchen and Food Preparation • Design Ventilation-Kitchen and Food Preparation • Design Max Occupant Density-Corridor • Design Ventilation-Corridor • Design Max Occupant Density-Laundry • Design Ventilation-Laundry • Design Max Occupant Density-All Others • Design Ventilation-All Others
11	Heating and Cooling System	**Based on Default from eQUEST**
12	Landscaping System	**No Option in eQUEST**

Table 3.3 Percent Area of Residential Low Rise

Activity Area Type	Detailed Items	% Area
1. General Living Space	Family/Den (300) + Living Room (300) + Bath#1 (40) + Bath#2 (40) + Bath-Master (70) + Closets (122) = 872	$\approx$35%
2. Bedroom	Bedroom#1 (180) + Bedroom#2 (180) + Bed-Master (252) = 612	$\approx$24%
3. Garage	Garage = 528	$\approx$21%
4. Dining Area	Dining Room = 208	$\approx$8%
5. Kitchen and Food Preparation	Kitchen (96) + Pantry (16) + Breakfast (40) = 152	$\approx$6%
6. Corridor	Hall = 64	$\approx$3%
7. Laundry	Laundry = 40	$\approx$2%
8. All Others	Entry = 24	$\approx$1%
Total: 2500 Square Feet		

Table 3.4 Two Settings for 46 Variables

Variable	Two Settings
Wall Construction	16 inch (2×4) Wood and 24 inch (2×4) Wood
Window-Glass Category	Double Clear/Tint and Double Low-e (e2 = 0.1)
Roof Construction	24 inch Wood and >24 inch Wood
Ground Floor Construction	2 inch Concrete and 4 inch Concrete
Ground Floor Interior Insulation	None and 1 inch Polystyrene
Ground Floor Cap	1.25 inch Concrete and 2 inch Concrete
Ground Floor Exterior/Cavity Insulation	None and 1 inch Polystyrene
Exterior Wall Finishes	Brick and Concrete
Exterior Wall Color	Light and Dark
Exterior Wall Insulation	None and 1 inch Polystyrene
Additional Wall Insulation	None and R-11 Batt
Interior Wall Insulation	None and 1 inch Polystyrene
Windows-Glass Type	1/8, 1/4 inch Clear and 1/8, 1/2 inch Clear
Windows-Frame Width	2 and 3
Window Height	4 and 6
Windows-Sill Height	2 and 3
Distance from Window-Overhangs	1 and 1.5
Exterior Roof Finish	Concrete and Built-up Roof
Exterior Roof Color	Light and Dark
Exterior Roof Insulation	None and 1 inch Polystyrene
Additional Roof Insulation	None and R-3 Batt
Ceiling Interior Finishes	Drywall Finish and Plaster Finish
Ceiling Batt Insulation	R-11 Batt and R-13 Batt
Pitched Roof	Standard Wood Framing and Advanced Wood Framing
Footprint X&Y	100×25 and 50×50
Orientation	N/S Component (Face North) and E/W Component (Face East)
Doors-Construction	Double Clear/Tint and Double Low-e (e2 = 0.1)
Door Glass Type	1/8, 1/4 inch Clear and 1/8, 1/2 inch Clear
Door Dimension-Height&Width	7,3 and 7,6
Door-Frame Width	2 and 3
Design Max Occupant Density-1	575 and 675
Design Ventilation-1	10 and 30
Design Max Occupant Density-2	575 and 675
Design Ventilation-2	10 and 30
Design Max Occupant Density-3	575 and 675
Design Ventilation-3	10 and 30
Design Max Occupant Density-4	5 and 105
Design Ventilation-4	10 and 30
Design Max Occupant Density-5	250 and 350
Design Ventilation-5	5 and 25
Design Max Occupant Density-6	100 and 200
Design Ventilation-6	5 and 25
Design Max Occupant Density-7	100 and 200
Design Ventilation-7	15 and 35
Design Max Occupant Density-8	100 and 200

Table 3.4 – *Continued*

Design Ventilation-8	5 and 25

Table 3.5 Random Variables

Variable	**Range**
Occupied Loads-Lighting-1	(0.5,2.5)
Occupied Loads-Task Lighting-1	(0.5,2.5)
Occupied Loads-Plug Loads-1	(0.5,2.5)
Unoccupied Loads-Occupancy-1	(0,2)
Unoccupied Loads-Lighting-1	(0,2)
Unoccupied Loads-Task Lighting-1	(0,2)
Unoccupied Loads-Plug Loads-1	(0,20)
Occupied Loads-Lighting-2	(0.5,2.5)
Occupied Loads-Task Lighting-2	(0.5,2.5)
Occupied Loads-Plug Loads-2	(0.5,2.5)
Unoccupied Loads-Occupancy-2	(0,2)
Unoccupied Loads-Lighting-2	(0,2)
Unoccupied Loads-Task Lighting-2	(0,2)
Unoccupied Loads-Plug Loads-2	(0,20)
Occupied Loads-Lighting-3	(0.5,2.5)
Occupied Loads-Task Lighting-3	(0.5,2.5)
Occupied Loads-Plug Loads-3	(0.5,2.5)
Unoccupied Loads-Occupancy-3	(0,2)
Unoccupied Loads-Lighting-3	(0,2)
Unoccupied Loads-Task Lighting-3	(0,2)
Unoccupied Loads-Plug Loads-3	(0,20)
Occupied Loads-Lighting-4	(0.5,2.5)
Occupied Loads-Task Lighting-4	(0.5,2.5)
Occupied Loads-Plug Loads-4	(0.5,2.5)
Unoccupied Loads-Occupancy-4	(0,2)
Unoccupied Loads-Lighting-4	(0,2)
Unoccupied Loads-Task Lighting-4	(0,2)
Unoccupied Loads-Plug Loads-4	(0,20)
Occupied Loads-Lighting-5	(0.5,2.5)
Occupied Loads-Task Lighting-5	(0.5,2.5)
Occupied Loads-Plug Loads-5	(0.5,2.5)
Unoccupied Loads-Occupancy-5	(0,2)
Unoccupied Loads-Lighting-5	(0,2)
Unoccupied Loads-Task Lighting-5	(0,2)
Unoccupied Loads-Plug Loads-5	(0,20)
Occupied Loads-Lighting-6	(0.5,2.5)
Occupied Loads-Task Lighting-6	(0.5,2.5)
Occupied Loads-Plug Loads-6	(0.5,2.5)
Unoccupied Loads-Occupancy-6	(0,2)
Unoccupied Loads-Lighting-6	(0,2)
Unoccupied Loads-Task Lighting-6	(0,2)
Unoccupied Loads-Plug Loads-6	(0,20)
Occupied Loads-Lighting-7	(0.5,2.5)
Occupied Loads-Task Lighting-7	(0.5,2.5)

Table 3.5 – *Continued*

Occupied Loads-Plug Loads-7	(0.5,2.5)
Unoccupied Loads-Occupancy-7	(0,2)
Unoccupied Loads-Lighting-7	(0,2)
Unoccupied Loads-Task Lighting-7	(0,2)
Unoccupied Loads-Plug Loads-7	(0,20)
Occupied Loads-Lighting-8	(0.5,2.5)
Occupied Loads-Task Lighting-8	(0.5,2.5)
Occupied Loads-Plug Loads-8	(0.5,2.5)
Unoccupied Loads-Occupancy-8	(0,2)
Unoccupied Loads-Lighting-8	(0,2)
Unoccupied Loads-Task Lighting-8	(0,2)
Unoccupied Loads-Plug Loads-8	(0,20)
Thermostat Cooling Setpoints-Occupied	(71,81)
Thermostat Cooling Setpoints-Unoccupied	(77,87)
Thermostat Heating Setpoints-Occupied	(65,75)
Thermostat Heating Setpoints-Unoccupied	(59,69)
Cooling Design Temperature-Indoor	(70,80)
Cooling Design Temperature-Supply	(50,60)
Heating Design Temperature-Indoor	(67,77)
Heating Design Temperature-Supply	(115,125)
Air Flows-Minimum Design Flow	(0,1)
HVAC-Economizer High Limit	(65,75)
Water Temp-Supply Water_NoResDomWH	(130,140)
Water Temp-Supply Water_ResDomWH	(105,115)

For the DACE screening analysis, this research employed an orthogonal array (OA) [65] experimental design that limited each factor variable to two settings (nominally the high and low of the ranges). An OA with 108 runs for up to 107 factors was selected from the R package: DoE.base [66]. Each run of the OA design specifies settings for the 46 factor variables for one run of eQUEST. In addition, for each run of eQUEST, one instance of the 68 random variables was sampled. The output from eQUEST was used to identify response variables for statistical analysis.

For main schedule information and the HVAC fan schedule for system #1, we assume everyday has the same schedule. The timeframe of building usage schedule is from 5 pm to 6 am every day. Other options are based on eQUEST default values. The first cost for a low-rise residential building was set as $250,000 with an annual maintenance cost of $1,800.

Finally, the remaining assumptions in eQUEST are: building type is Unknown, Custom or Mixed Use since there was the only option for a residential low rise building. Based on eQUEST default settings, code analysis is None, jurisdiction is CA Title24, location set is California, region is Los Angeles Area, city is Los Angeles AP, electric is SCE (CA), rate is GS-1 (3) (< 20 kW, three-phase service), and gas is SCG (CA) and rate is GN-10 (buildings with < 20800 therms/mo). This building simulation with 108 runs was completed in 2010, so Analysis Year in eQUEST is 2010. Ground floor exposure is Over Crawl Space since crawl space foundation has some advantages over slab foundation, for example, it is easy to install wiring and plumbing for crawl space foundation [67-68]. Moreover, ground floor finish is Carpet (No Pad); windows frame type is a fixed Aluminum w/o Brick. No daylighting control is considered for this residential building case study, footprint shape is Rectangle, zoning pattern is One per Floor; exterior door types is Glass and Opaque (Wood, Solid Core Flush, 1-3/8 inch for the garage door, height 7 feet (ft) and width 16 ft), #Doors by Orientation is 2 (front door and back door), and doors frame type is Aluminum w/o Brick. Windows area specification method is Percent of Gross Wall Area, and typical window width is 3 ft. Window percent glass is assumed to first be 25% on all four sides. Since eQUEST Wizard always places doors at the center, the front door and garage door are adhered together on the front (North or East) side, this research used a custom window option. Three windows are assumed between the front door and the garage door. The three windows of X axes are 17 ft, 21 ft and 25 ft. The front side of footprint dimension X axes are 100 ft and 50 ft, and floor-to-floor height is 12 ft. When window area that is window height of 6 ft multiplied by typical window width of 3 ft is 18 ft^2, and there are 11 windows, the front window side area is 100 ft by 12 ft. When window area that is window height of 4 ft multiplied by typical window width of 3 ft is 12 ft^2, and there are 15 windows, the front side window area is 100 ft by 12 ft. When window area that is window height of 6 ft multiplied by typical window width of 3 ft is 18 ft^2, and there are 7 windows, the front side window area is 50 ft by 12 ft. When window area that is window height of 4 ft multiplied by typical window width of 3 ft is 12 ft^2, and there are 7 windows, the front side window area is 50 ft by 12 ft. Therefore, the four actual window percent glass measurements for both the North and East side are equal at

16.5% (i.e. $\frac{18\times11}{100\times12}$), 15% (i.e. $\frac{12\times15}{100\times12}$), 21% (i.e. $\frac{18\times7}{50\times12}$), and 14% (i.e. $\frac{12\times7}{50\times12}$). Thus, only the actual value of the designated front side is not 25%. For example, the first eQUEST run had the front facing East, so window percent glass factor variables were %Window-North (25%), %Window-South (25%), %Window-East (16.5%) and %Window-West (25%). In addition, shade depths are all 1 ft, Interior finish is drywall finish, Batt insulation is R-30, Rigid insulation is 1 1/2 inch polystyrene (R-6). For performance evaluation, measure type Whole Site/Building is selected from EEM Wizard.

Four response variables (performance metrics) were selected from the eQUEST output: annual source energy-total million British thermal unit (Mbtu) (Y_1), HVAC energy-total Mbtu (Y_2), annual utility cost-total (Y_3) and LCC (Y_4). For example, the output values of 108 runs with pitched roof are shown in Figure 3.1-3.4. The multivariate analysis of these performance metrics will guide the development of multi-response models that will be incorporated into an MSMO framework to evaluate how building decisions affect the multiple performance objectives.

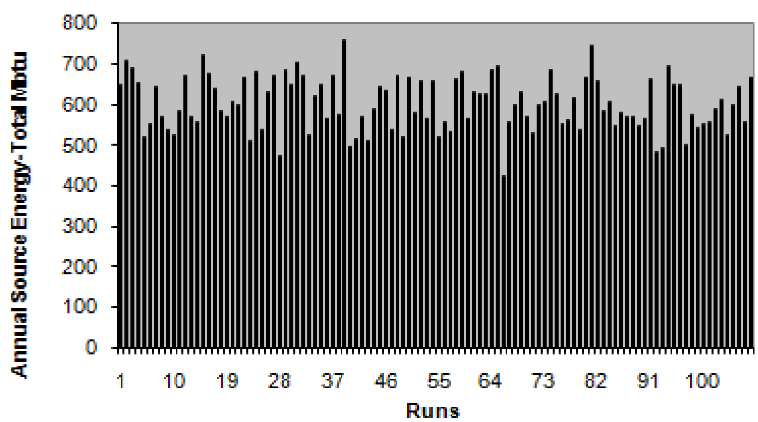


Figure 3.1 Annual Source Energy-Total Mbtu (Y_1)

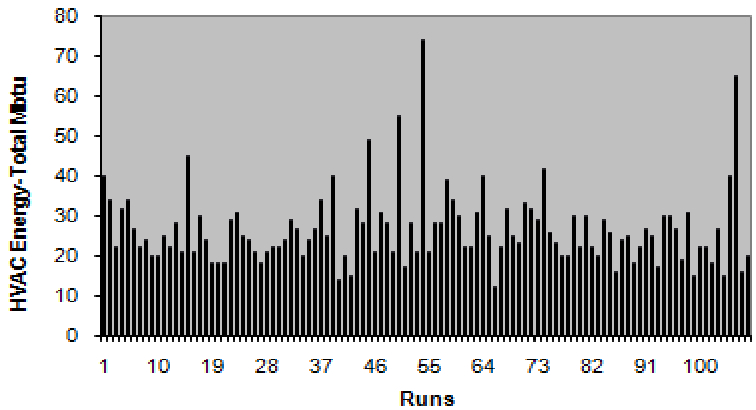


Figure 3.2 HVAC Energy-Total Mbtu (Y_2)

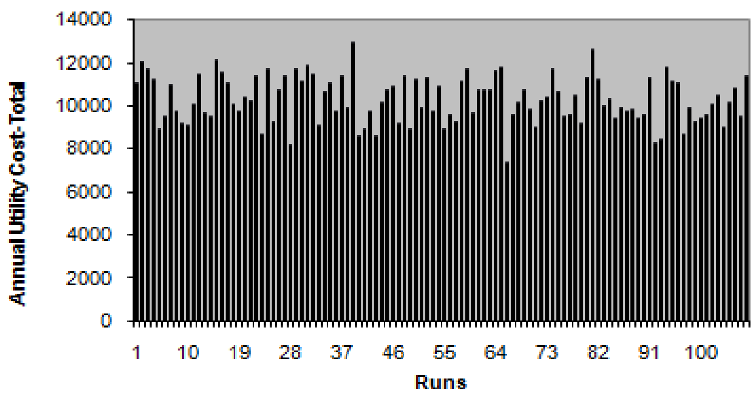


Figure 3.3 Annual Utility Cost-Total (Y_3)

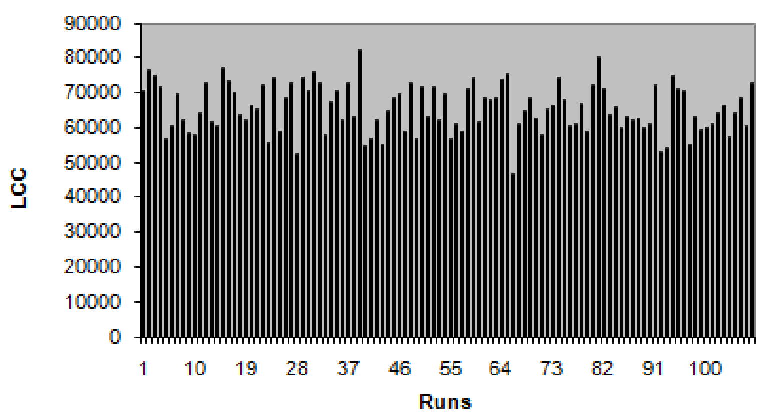


Figure 3.4 LCC (Y_4)

3.1.1.2 Multivariate Analysis

Multivariate analysis of variance (MANOVA) [21] is used to study the effects of factors on multiple response variables. MANOVA is particularly appropriate when the response variables are correlated. Using SAS software [69], both ANOVA on individual response variables (Y_1, Y_2, Y_3, Y_4) and MANOVA were conducted. MANOVA can be used to calculate *p*-values for the four performance metrics (response variables).

The *p*-values from the hypothesis tests on the factor effects are shown in Table 3.6 (108 runs) for the factors that were statistically significant at the 0.10 level (i.e., *p*-value < 0.10) for at least one of the analyses. These 13 factor variables identified were: Ground Floor Exterior/Cavity Insulation, Exterior Wall Finishes, Interior Wall Insulation, Windows-Sill Height, Distance from Window-Overhangs, Additional Roof Insulation, Ceiling Interior Finishes, Footprint X&Y, Orientation, Design Max Occupant Density- Residential (General Living Space), Design Max Occupant Density-Dining Area, Design Ventilation-Dining Area, and Design Max Occupant Density-Kitchen and Food Preparation.

The previous method utilized an OA with 108 runs assuming a pitched roof, and then this research additionally conducted 108 runs without a pitched roof. Thus, the combined set had 216 runs that were analyzed in this research [70-71]. The building simulation with 216 runs was completed in 2010, so Analysis Year in eQUEST is 2010. Based on the results in Table 3.7, factors that may affect the four green building performance metrics were identified as having at least one significant *p*-value (< 0.05) in the row. The following 32 factors were identified: Footprint X&Y, Orientation, Ground Floor Interior Insulation, Ground Floor Cap, Ground Floor Exterior/Cavity Insulation, Wall Construction, Exterior Wall Finishes, Exterior Wall Color, Exterior Wall Insulation, Additional Wall Insulation, Interior Wall Insulation, Window-Glass Category, Window Height, Windows-Sill Height, Distance from Window-Overhangs, Doors Construction, Door Dimension-Height&Width, Roof Construction, Exterior Roof Finish, Exterior Roof Color, Additional Roof Insulation, Ceiling Interior Finishes, Ceiling Batt Insulation, Design

Max Occupant Density-Residential (General Living Space), Design Max Occupant Density-Residential (Bedroom), Design Ventilation-Residential (Bedroom), Design Ventilation-Residential (Garage), Design Max Occupant Density-Dining Area, Design Ventilation-Dining Area, Design Max Occupant Density-Kitchen and Food Preparation, Design Ventilation-Corridor, Design Max Occupant Density-Laundry.

3.1.2 Comparison of Several Analyses

In Tables 3.8-3.10, this research compared the ANOVA on individual response variables and a multivariate ANOVA from SAS in Section 3.1.1 with multivariate regression tree splitting using mvpart from R, the MD method, which converts multiple responses into a single response for ANOVA, and the software CART and MARS from Salford Systems [72]. Almost all 46 factors were important in at least one analysis. Thus, this research proceeded with all 46 factors for the additional study in the next chapter. The further eQUEST ideal settings for 46 factors are discussed in chapter 4.

Table 3.6 Stages and *P* Values for 108 Runs

Stage	Building Category	46 Factors	*P* Values				
			Y_1	Y_2	Y_3	Y_4	MANOVA
1	Siting Options	• Footprint X&Y	0.8584	**0.0159**	0.8391	0.8409	**0.0543**
		• Orientation	0.2060	**0.0095**	0.2245	0.2238	**0.0296**
2	Electrical System	**Based on Default from eQUEST**					
3	Wells and Septic System	**No Option in eQUEST**					
4	Foundation System	• Ground Floor Construction	0.7967	0.5807	0.7657	0.7663	0.6856
		• Ground Floor Interior Insulation	0.8085	0.8925	0.8364	0.8369	0.4814
		• Ground Floor Cap	0.5002	0.4693	0.5606	0.5597	0.2358
		• Ground Floor Exterior/Cavity Insulation	0.2585	**0.0292**	0.2736	0.2727	0.2381
5	Plumbing System	**No Option in eQUEST**					
6	Wall System	• Wall Construction	0.2028	0.5155	0.2048	0.2045	0.6634
		• Exterior Wall Finishes	**0.0271**	**0.0436**	**0.0229**	**0.0229**	**0.0147**
		• Exterior Wall Color	0.4755	0.3844	0.4234	0.4258	0.1956
		• Exterior Wall Insulation	0.3333	0.1609	0.3662	0.3650	0.5977
		• Additional Wall Insulation	0.5557	0.1222	0.6231	0.6208	0.4183
		• Interior Wall Insulation	**0.0149**	0.5155	**0.0119**	**0.0119**	**0.0570**
7	Window System	• Window-Glass Category	0.3112	0.9706	0.2758	0.2769	0.4031
		• Windows-Glass Type	0.7089	0.3978	0.4445	0.4457	0.7568
		• Windows-Frame Width	0.9939	0.5315	0.9807	0.9804	0.8877
		• Window Height	0.5099	0.5315	0.5176	0.5161	0.2304
		• Windows-Sill Height	**0.0578**	0.4399	**0.0586**	**0.0585**	0.4611
		• Distance from Window-Overhangs	**0.0215**	0.9119	**0.0178**	**0.0180**	**0.0400**
8	Door System	• Doors Construction	0.1344	0.7034	0.1258	0.1263	0.3950
		• Door Glass Type	0.2573	0.4399	0.2519	0.2521	0.6864
		• Door Dimension-Height&Width	0.4425	0.7034	0.4962	0.4957	0.1916
		• Door-Frame Width	0.7157	0.8154	0.7128	0.7127	0.9964
9	Roof System	• Roof Construction	0.7089	0.2265	0.7842	0.7809	0.2516
		• Exterior Roof Finish	0.6999	0.4399	0.8028	0.8002	0.1166
		• Exterior Roof Color	0.3564	0.3335	0.4112	0.4100	0.2745
		• Exterior Roof Insulation	0.9308	0.7775	0.9538	0.9542	0.6998
		• Additional Roof Insulation	0.4335	0.8154	0.4375	0.4364	**0.0628**
		• Ceiling Interior Finishes	0.1000	0.2764	**0.0869**	**0.0870**	0.1455

Table 3.6 – *Continued*

		• Ceiling Batt Insulation • Pitched Roof	0.1834 0.6752	0.2456 0.4545	0.2001 0.6004	0.1993 0.6029	0.5787 0.3199
10	Ventilation System	• Design Max Occupant Density-Residential (General Living Space)	**0.0790**	**0.0192**	**0.0941**	**0.0934**	0.1575
		• Design Ventilation-Residential (General Living Space)	0.3333	0.5976	0.3424	0.3427	0.6002
		• Design Max Occupant Density-Residential (Bedroom)	0.4177	0.8345	0.4350	0.4340	0.4683
		• Design Ventilation- Residential (Bedroom)	0.7134	0.6673	0.7782	0.7766	0.5164
		• Design Max Occupant Density-Residential (Garage)	0.5177	0.2764	0.5224	0.5219	0.8078
		• Design Ventilation-Residential (Garage)	0.4533	0.9119	0.4411	0.4424	0.3646
		• Design Max Occupant Density-Dining Area	0.5376	**<0.0001**	0.6082	0.6019	**<0.0001**
		• Design Ventilation-Dining Area	0.7780	**0.0005**	0.6209	0.6273	**0.0009**
		• Design Max Occupant Density-Kitchen and Food Preparation	**0.0469**	0.1166	**0.0504**	**0.0504**	**0.0530**
		• Design Ventilation-Kitchen and Food Preparation	0.3084	0.3583	0.3050	0.3047	0.7686
		• Design Max Occupant Density-Corridor	0.4533	0.5155	0.4559	0.4566	0.6962
		• Design Ventilation-Corridor	0.3170	0.9119	0.2789	0.2796	0.3771
		• Design Max Occupant Density-Laundry	0.5177	0.4256	0.5236	0.5246	0.5131
		• Design Ventilation-Laundry	0.7897	0.3583	0.8497	0.8481	0.6610
		• Design Max Occupant Density-All Others	0.3692	0.9314	0.3715	0.3720	0.7634
		• Design Ventilation-All Others	0.3888	0.7587	0.3955	0.3953	0.6399
11	Heating and Cooling System	**Based on Default from eQUEST**					
12	Landscaping System	**No Option in eQUEST**					

Table 3.7 Stages and *P* Values for 216 Runs

Stage	Building Category	46 Factors	*P* Values				
			Y_1	Y_2	Y_3	Y_4	MANOVA
1	Siting Options	• Footprint X&Y	0.7702	**<0.0001**	0.7380	0.7410	**<0.0001**
		• Orientation	**0.0351**	**<0.0001**	**0.0442**	**0.0439**	**<0.0001**
2	Electrical System	**Based on Default from eQUEST**					
3	Wells and Septic System	**No Option in eQUEST**					
4	Foundation System	• Ground Floor Construction	0.6714	0.3474	0.6221	0.6230	0.1662
		• Ground Floor Interior Insulation	0.6788	0.7793	0.7256	0.7265	**0.0271**
		• Ground Floor Cap	0.2623	0.2466	0.3327	0.3317	**0.0027**
		• Ground Floor Exterior/Cavity Insulation	0.0614	**0.0003**	0.0695	0.0690	**0.0035**
5	Plumbing System	**No Option in eQUEST**					
6	Wall System	• Wall Construction	**0.0355**	0.3683	**0.0356**	**0.0354**	0.1110
		• Exterior Wall Finishes	**0.0002**	**0.0011**	**0.0002**	**0.0002**	**<0.0001**
		• Exterior Wall Color	0.2344	0.1452	0.1820	0.1843	**0.0037**
		• Exterior Wall Insulation	0.1069	**0.0147**	0.1324	0.1314	0.0922
		• Additional Wall Insulation	0.3212	**0.0112**	0.4048	0.4018	**0.0464**
		• Interior Wall Insulation	**<0.0001**	0.2895	**<0.0001**	**<0.0001**	**<0.0001**
7	Window System	• Window-Glass Category	0.0905	1.0000	0.0690	0.0696	**0.0316**
		• Windows-Glass Type	0.2255	0.1398	0.2049	0.2062	0.2100
		• Windows-Frame Width	0.9939	0.3474	0.9765	0.9763	0.4733
		• Window Height	0.2634	0.2632	0.2747	0.2728	**0.0028**
		• Windows-Sill Height	**0.0015**	0.2083	**0.0015**	**0.0015**	**0.0376**
		• Distance from Window-Overhangs	**0.0001**	0.9203	**<0.0001**	**<0.0001**	**<0.0001**
8	Door System	• Doors Construction	**0.0126**	0.5618	**0.0106**	**0.0107**	**0.0249**
		• Door Glass Type	0.0584	0.1746	0.0551	0.0552	0.1580
		• Door Dimension-Height&Width	0.2014	0.4964	0.2602	0.2595	**0.0013**
		• Door-Frame Width	0.5523	0.7488	0.5453	0.5453	0.9734
9	Roof System	• Roof Construction	0.5288	0.0511	0.6523	0.6471	**0.0061**
		• Exterior Roof Finish	0.5222	0.1565	0.6897	0.6853	**0.0004**
		• Exterior Roof Color	0.1119	0.0980	0.1532	0.1522	**0.0113**
		• Exterior Roof Insulation	0.8929	0.5221	0.9283	0.9294	0.1180
		• Additional Roof Insulation	0.1909	0.6169	0.1942	0.1931	**<0.0001**
		• Ceiling Interior Finishes	**0.0059**	0.0612	**0.0042**	**0.0042**	**0.0008**

Table 3.7 – *Continued*

		• Ceiling Batt Insulation • Pitched Roof	**0.0276** 0.8909	0.0612 0.5092	**0.0339** 0.8964	**0.0336** 0.8957	0.1153 0.9627
10	Ventilation System	• Design Max Occupant Density-Residential (General Living Space)	**0.0034**	**0.0001**	**0.0053**	**0.0052**	**0.0013**
		• Design Ventilation-Residential (General Living Space)	0.1069	0.3683	0.1128	0.1130	0.1030
		• Design Max Occupant Density-Residential (Bedroom)	0.1760	0.7038	0.1954	0.1943	**0.0353**
		• Design Ventilation- Residential (Bedroom)	0.5322	0.4839	0.6397	0.6373	**0.0317**
		• Design Max Occupant Density-Residential (Garage)	0.2788	0.0730	0.2851	0.2844	0.3366
		• Design Ventilation-Residential (Garage)	0.2207	0.7337	0.2045	0.2060	**0.0147**
		• Design Max Occupant Density-Dining Area	0.2901	**<0.0001**	0.3875	0.3788	**<0.0001**
		• Design Ventilation-Dining Area	0.6659	**<0.0001**	0.4223	0.4316	**<0.0001**
		• Design Max Occupant Density-Kitchen and Food Preparation	**0.0010**	**0.0125**	**0.0012**	**0.0012**	**<0.0001**
		• Design Ventilation-Kitchen and Food Preparation	0.0882	0.1508	0.0880	0.0877	0.3515
		• Design Max Occupant Density-Corridor	0.2114	0.2466	0.2127	0.2134	0.1757
		• Design Ventilation-Corridor	0.0939	0.8102	0.0716	0.0720	**0.0406**
		• Design Max Occupant Density-Laundry	0.2833	0.1684	0.2923	0.2936	**0.0276**
		• Design Ventilation-Laundry	0.6677	0.1346	0.7577	0.7550	0.2804
		• Design Max Occupant Density-All Others	0.1319	0.8886	0.1376	0.1379	0.2074
		• Design Ventilation-All Others	0.1446	0.6597	0.1522	0.1520	0.1266
11	Heating and Cooling System	**Based on Default from eQUEST**					
12	Landscaping System	**No Option in eQUEST**					

Table 3.8 Six Analysis Methods with 216 Runs (Alpha Level 0.01)

Individual analyses and MANOVA	**MD-Single Y Individual analyses and MANOVA**	**MD-Single Y MARS-Salford**	**MD-Single Y CART-Salford**	**MVPART-RPART Code**	**MVPART-MVPART Code**
				Wall Construction	
Roof Construction			Roof Construction	Roof Construction	Roof Construction
				Ground Floor Construction	
				Ground Floor Interior Insulation	Ground Floor Interior Insulation
Ground Floor Cap					
Ground Floor Exterior/Cavity Insulation					
Exterior Wall Finishes	Exterior Wall Finishes	Exterior Wall Finishes		Exterior Wall Finishes	Exterior Wall Finishes
Exterior Wall Color					
				Additional Wall Insulation	Additional Wall Insulation
Interior Wall Insulation	Interior Wall Insulation	Interior Wall Insulation		Interior Wall Insulation	Interior Wall Insulation
				Windows-Glass Type	Windows-Glass Type
				Windows-Frame Width	Windows-Frame Width
Window Height				Window Height	Window Height
Windows-Sill Height					
Distance from Window-Overhangs				Distance from Window-Overhangs	Distance from Window-Overhangs
Exterior Roof Finish				Exterior Roof Finish	Exterior Roof Finish
				Exterior Roof Insulation	Exterior Roof Insulation
Additional Roof Insulation					

Table 3.8 – *Continued*

Ceiling Interior Finishes				Ceiling Interior Finishes	Ceiling Interior Finishes
	Ceiling Batt Insulation	Ceiling Batt Insulation	Ceiling Batt Insulation	Ceiling Batt Insulation	
Footprint X&Y				Footprint X&Y	Footprint X&Y
Orientation					
				Doors-Construction	Doors-Construction
				Doors-Glass Type	Doors-Glass Type
Door Dimension-Height&Width					
Design Max Occupant Density-Residential (General Living Space)			Design Max Occupant Density-Residential (General Living Space)		
				Design Ventilation-Residential (General Living Space)	Design Ventilation-Residential (General Living Space)
				Design Max Occupant Density-Residential (Garage)	Design Max Occupant Density-Residential (Garage)
Design Max Occupant Density-Dining Area	Design Max Occupant Density-Dining Area	Design Max Occupant Density-Dining Area	Design Max Occupant Density-Dining Area	Design Max Occupant Density-Dining Area	Design Max Occupant Density-Dining Area
Design Ventilation-Dining Area	Design Ventilation-Dining Area	Design Ventilation-Dining Area	Design Ventilation-Dining Area	Design Ventilation-Dining Area	Design Ventilation-Dining Area
Design Max Occupant Density-Kitchen and Food Preparation				Design Max Occupant Density-Kitchen and Food Preparation	Design Max Occupant Density-Kitchen and Food Preparation
				Design Ventilation-Laundry	Design Ventilation-Laundry

Table 3.9 Six Analysis Methods with 216 Runs (Alpha Level 0.05)

Individual analyses and MANOVA	**MD-Single Y Individual analyses and MANOVA**	**MD-Single Y MARS-Salford**	**MD-Single Y CART-Salford**	**MVPART-RPART Code**	**MVPART-MVPART Code**
Wall Construction				Wall Construction	
Windows-Glass Category					
Roof Construction			Roof Construction	Roof Construction	Roof Construction
				Ground Floor Construction	
Ground Floor Interior Insulation				Ground Floor Interior Insulation	Ground Floor Interior Insulation
Ground Floor Cap					
Ground Floor Exterior/Cavity Insulation					
Exterior Wall Finishes	Exterior Wall Finishes	Exterior Wall Finishes		Exterior Wall Finishes	Exterior Wall Finishes
Exterior Wall Color					
Exterior Wall Insulation					
Additional Wall Insulation				Additional Wall Insulation	Additional Wall Insulation
Interior Wall Insulation	Interior Wall Insulation	Interior Wall Insulation		Interior Wall Insulation	Interior Wall Insulation
	Windows-Glass Type			Windows-Glass Type	Windows-Glass Type
				Windows-Frame Width	Windows-Frame Width
Window Height				Window Height	Window Height
Windows-Sill Height					
Distance from Window-Overhangs				Distance from Window-Overhangs	Distance from Window-Overhangs

Table 3.9 – *Continued*

Exterior Roof Finish				Exterior Roof Finish	Exterior Roof Finish
Exterior Roof Color					
				Exterior Roof Insulation	Exterior Roof Insulation
Additional Roof Insulation					
Ceiling Interior Finishes				Ceiling Interior Finishes	Ceiling Interior Finishes
Ceiling Batt Insulation	Ceiling Batt Insulation	Ceiling Batt Insulation	Ceiling Batt Insulation	Ceiling Batt Insulation	
Footprint X&Y				Footprint X&Y	Footprint X&Y
Orientation					
Doors Construction				Doors-Construction	Doors-Construction
				Doors-Glass Type	Doors-Glass Type
Door Dimension-Height&Width					
Design Max Occupant Density-Residential (General Living Space)			Design Max Occupant Density-Residential (General Living Space)		
				Design Ventilation-Residential (General Living Space)	Design Ventilation-Residential (General Living Space)
Design Max Occupant Density-Residential (Bedroom)					
Design Ventilation-Residential (Bedroom)					
	Design Max Occupant Density-Residential (Garage)			Design Max Occupant Density-Residential (Garage)	Design Max Occupant Density-Residential (Garage)

Table 3.9 – *Continued*

Design Ventilation-Residential (Garage)					
Design Max Occupant Density-Dining Area	Design Max Occupant Density-Dining Area	Design Max Occupant Density-Dining Area	Design Max Occupant Density-Dining Area	Design Max Occupant Density-Dining Area	Design Max Occupant Density-Dining Area
Design Ventilation-Dining Area	Design Ventilation-Dining Area	Design Ventilation-Dining Area	Design Ventilation-Dining Area	Design Ventilation-Dining Area	Design Ventilation-Dining Area
Design Max Occupant Density-Kitchen and Food Preparation				Design Max Occupant Density-Kitchen and Food Preparation	Design Max Occupant Density-Kitchen and Food Preparation
Design Ventilation-Corridor					
Design Max Occupant Density-Laundry					
				Design Ventilation-Laundry	Design Ventilation-Laundry

Table 3.10 Six Analysis Methods with 216 Runs (Alpha Level 0.10)

Individual analyses and MANOVA	**MD-Single Y Individual analyses and MANOVA**	**MD-Single Y MARS-Salford**	**MD-Single Y CART-Salford**	**MVPART-RPART Code**	**MVPART-MVPART Code**
Wall Construction				Wall Construction	
Windows-Glass Category					
Roof Construction			Roof Construction	Roof Construction	Roof Construction
				Ground Floor Construction	
Ground Floor Interior Insulation				Ground Floor Interior Insulation	Ground Floor Interior Insulation
Ground Floor Cap					

Table 3.10 – *Continued*

Ground Floor Exterior/Cavity Insulation					
Exterior Wall Finishes	Exterior Wall Finishes	Exterior Wall Finishes		Exterior Wall Finishes	Exterior Wall Finishes
Exterior Wall Color					
Exterior Wall Insulation					
Additional Wall Insulation				Additional Wall Insulation	Additional Wall Insulation
Interior Wall Insulation	Interior Wall Insulation	Interior Wall Insulation		Interior Wall Insulation	Interior Wall Insulation
	Windows-Glass Type			Windows-Glass Type	Windows-Glass Type
				Windows-Frame Width	Windows-Frame Width
Window Height				Window Height	Window Height
Windows-Sill Height					
Distance from Window-Overhangs				Distance from Window-Overhangs	Distance from Window-Overhangs
Exterior Roof Finish				Exterior Roof Finish	Exterior Roof Finish
Exterior Roof Color	Exterior Roof Color				
				Exterior Roof Insulation	Exterior Roof Insulation
Additional Roof Insulation					
Ceiling Interior Finishes				Ceiling Interior Finishes	Ceiling Interior Finishes
Ceiling Batt Insulation	Ceiling Batt Insulation	Ceiling Batt Insulation	Ceiling Batt Insulation	Ceiling Batt Insulation	
Footprint X&Y				Footprint X&Y	Footprint X&Y
Orientation					
Doors Construction				Doors-Construction	Doors-Construction
Door Glass Type				Doors-Glass Type	Doors-Glass Type

Table 3.10 – *Continued*

Door Dimension-Height&Width					
Design Max Occupant Density-Residential (General Living Space)	Design Max Occupant Density-Residential (General Living Space)		Design Max Occupant Density-Residential (General Living Space)		
				Design Ventilation-Residential (General Living Space)	Design Ventilation-Residential (General Living Space)
Design Max Occupant Density-Residential (Bedroom)					
Design Ventilation-Residential (Bedroom)					
Design Max Occupant Density-Residential (Garage)	Design Max Occupant Density-Residential (Garage)			Design Max Occupant Density-Residential (Garage)	Design Max Occupant Density-Residential (Garage)
Design Ventilation-Residential (Garage)					
Design Max Occupant Density-Dining Area	Design Max Occupant Density-Dining Area	Design Max Occupant Density-Dining Area	Design Max Occupant Density-Dining Area	Design Max Occupant Density-Dining Area	Design Max Occupant Density-Dining Area
Design Ventilation-Dining Area	Design Ventilation-Dining Area	Design Ventilation-Dining Area	Design Ventilation-Dining Area	Design Ventilation-Dining Area	Design Ventilation-Dining Area
Design Max Occupant Density-Kitchen and Food Preparation	Design Max Occupant Density-Kitchen and Food Preparation			Design Max Occupant Density-Kitchen and Food Preparation	Design Max Occupant Density-Kitchen and Food Preparation
Design Ventilation-Kitchen and Food					

Table 3.10 – *Continued*

Preparation					
Design Ventilation-Corridor	Design Ventilation-Corridor				
Design Max Occupant Density-Laundry					
				Design Ventilation-Laundry	Design Ventilation-Laundry

CHAPTER 4

MULTI-RESPONSE TREE-BASED MODELS USING THE METHOD OF SEEMINGLY UNRELATED REGRESSIONS

4.1 Second Design of Experiments

This research modified some previous assumptions and factor variables. For example, most residential buildings do not have window overhangs, so Distance from Window-Overhangs is not considered to be a factor variable, and shade depths are not represented in the assumptions. Each of the four sides of the building has a percent glass variable that is considered to be important, and custom window options cannot be used. For the main schedule information for building operations, the timeframe is from 5pm to 8am on the weekday with occupancy percent 90%, ambient lighting load percent 90%, and equipment load percent 90%. On the weekend and holidays, the timeframe is all day with occupancy percent 50%, ambient lighting load percent 50%, and equipment load percent 50%. This building simulation with 192 runs was completed in 2012, so Analysis Year in eQUEST is 2012. Others are the same as the previous assumptions in Section 3.1.1.1.

For this second experimental design, the 46 variables are described by three types: discrete-numerical, discrete-categorical, and continuous. In Table 4.1, there are 10 factors with two levels, one factor with three levels, and 19 factors with four levels. To accommodate this combination of factors, the second design employed a mixed-level orthogonal array $2^{19}3^{1}4^{23}$ design of strength 2 (DoE.base [66]) that allows 19 factors with 2 levels, one with 3 levels, and 23 with 4 levels. There are 16 factors that are discrete-numerical type, but only limited 4 levels are permitted in eQUEST. Six variables have 4 different levels for inch concretes and polystyrenes. There are three variables that have R-values [73-76] using U.S. units for batt insulation. The R-value which measures thermal resistance through a given thickness of material is

$$R = \frac{\Delta T}{Q}, \tag{4.1}$$

where ΔT is temperature difference, and Q is heat flux. The unit of R-value is ft^2·°F·hour/Btu. In R-values, the integer numbers are standard insulation types. For example, R-3 means R-value per inch of thickness is 3. When the thickness is changed, the R-values are varies. Thus, the R-values are numerical. Since the R-value is additive, the total value 3 can be cumulated from different R-values of insulation materials. Other values of variables in discrete-numerical types have limited specific ranges. There are 14 variables that are discrete-categorical types, where 10 each have 2 levels, one has 3 levels, and 3 have 4 levels. Interior wall insulation, which is considered a discrete-categorical type, only has two options: none, and 1 inch. Moreover, this research also employed a Sobol' low-discrepancy sequence [77] since eQUEST allows 8 continuous factors for maximum occupant density and 8 continuous factors for ventilation. In other words, there are 16 factors that are continuous and can be treated continuously in eQUEST. These ideal settings are shown in Table 4.1. This research also identified 68 uncontrollable variables and sampled the values from uniform distributions.

Table 4.1 eQUEST Ideal Settings for Training

Variables	Settings	Types
Ground Floor Construction (x_1)	• 2 inch Concrete • 4 inch Concrete • 6 inch Concrete • 8 inch Concrete	Discrete-Numerical
Ground Floor Interior Insulation (x_2)	• 1 inch Polystyrene • 1 1/2 inch Polystyrene • 2 inch Polystyrene • 3 inch Polystyrene	Discrete-Numerical
Ground Floor Cap (x_3)	• 1.25 inch Lightweight Concrete • 2 inch Lightweight Concrete • 3 inch Lightweight Concrete • 4 inch Lightweight Concrete	Discrete-Numerical
Ground Floor Exterior/Cavity Insulation (x_4)	• 1 inch Polystyrene • 2 inch Polystyrene • 3 inch Polystyrene	Discrete-Numerical

Table 4.1 – *Continued*

	• 4 inch Polystyrene	
Exterior Wall Insulation (x_5)	• 1 inch Polystyrene • 1 1/2 inch Polystyrene • 2 inch Polystyrene • 3 inch Polystyrene	Discrete-Numerical
Additional Wall Insulation (x_6)	• R-3 Batt • R-7 Batt • R-11 Batt • R-13 Batt	Discrete-Numerical
%Window-North (x_7)	• 10 • 15 • 20 • 25	Discrete-Numerical
%Window-South (x_8)	• 10 • 15 • 20 • 25	Discrete-Numerical
%Window-East (x_9)	• 10 • 15 • 20 • 25	Discrete-Numerical
%Window-West (x_{10})	• 10 • 15 • 20 • 25	Discrete-Numerical
Additional Roof Insulation (x_{11})	• R-7 Batt • R-19 Batt • R-30 Batt • R-49 Batt	Discrete-Numerical
Ceiling Batt Insulation (x_{12})	• R-13 Batt • R-19 Batt • R-21 Batt • R-30 Batt	Discrete-Numerical
Exterior Roof Insulation (x_{13})	• 1 inch Polystyrene • 1 1/2 inch Polystyrene • 2 inch Polystyrene • 3 inch Polystyrene	Discrete-Numerical
Footprint X (x_{14})	• 100 • 70.7 • 62.5 • 50	Discrete-Numerical
Door Dimension-Width (x_{15})	• 3 • 4 • 5 • 6	Discrete-Numerical
Door-Frame Width (x_{16})	• 2 • 2.3 • 2.7 • 3	Discrete-Numerical

Table 4.1 – *Continued*

Design Max Occupant Density-Residential (General Living Space) (x_{17})	Range: 575 to 675	Continuous
Design Ventilation-Residential (General Living Space) (x_{18})	Range: 10 to 30	Continuous
Design Max Occupant Density-Residential (Bedroom) (x_{19})	Range: 575 to 675	Continuous
Design Ventilation-Residential (Bedroom) (x_{20})	Range: 10 to 30	Continuous
Design Max Occupant Density-Residential (Garage) (x_{21})	Range: 575 to 675	Continuous
Design Ventilation-Residential (Garage) (x_{22})	Range: 10 to 30	Continuous
Design Max Occupant Density-Dining Area (x_{23})	Range: 5 to 105	Continuous
Design Ventilation-Dining Area (x_{24})	Range: 10 to 30	Continuous
Design Max Occupant Density-Kitchen and Food Preparation (x_{25})	Range: 250 to 350	Continuous
Design Ventilation-Kitchen and Food Preparation (x_{26})	Range: 5 to 25	Continuous
Design Max Occupant Density-Corridor (x_{27})	Range: 100 to 200	Continuous
Design Ventilation-Corridor (x_{28})	Range: 5 to 25	Continuous
Design Max Occupant Density-Laundry (x_{29})	Range: 100 to 200	Continuous
Design Ventilation-Laundry (x_{30})	Range: 15 to 35	Continuous
Design Max Occupant Density-All Others (x_{31})	Range: 100 to 200	Continuous
Design Ventilation-All Others (x_{32})	Range: 5 to 25	Continuous
Wall Construction (x_{33})	• Wood Frame, 2×4, 16 inch o.c. (a) • Wood Frame, 2×4, 24 inch o.c. (b)	Discrete-Categorical
Windows-Glass Category (x_{34})	• Double Clear/Tint (a) • Double Low-e (e2 = 0.1) (b)	Discrete-Categorical
Roof Construction (x_{35})	• Wood Advanced Frame, 24 inch o.c. (a) • Wood Advanced Frame, >24 inch o.c. (b)	Discrete-Categorical
Exterior Wall Finishes (x_{36})	• Brick (a) • Concrete (b)	Discrete-Categorical
Exterior Wall Color (x_{37})	• Light (a) • Dark (b)	Discrete-Categorical

Table 4.1 – *Continued*

Interior Wall Insulation (x_{38})	• None (a) • 1 inch Polystyrene (b)	Discrete-Categorical
Exterior Roof Finish (x_{39})	• Concrete (a) • Built-up Roof (b)	Discrete-Categorical
Exterior Roof Color (x_{40})	• Light (a) • Dark (b)	Discrete-Categorical
Doors-Construction (x_{41})	• Double Clear/Tint (a) • Double Low-e (e2 = 0.1) (b)	Discrete-Categorical
Pitched Roof (x_{42})	• Without Pitched Roof (a) • With Pitched Roof (b)	Discrete-Categorical
Ceiling Interior Finishes (x_{43})	• Lay-In Acoustic Tile (a) • Drywall Finish (b) • Plaster Finish (c)	Discrete-Categorical
Windows-Glass Type (x_{44})	• Clear 1/8, 1/4 inch Air (a) • Clear 1/8, 1/2 inch Air (b) • Clear 1/4, 1/4 inch Air (c) • Clear 1/4, 1/2 inch Air (d)	Discrete-Categorical
Orientation (x_{45})	• N/S Component (Face North) (a) • N/S Component (Face South) (b) • E/W Component (Face East) (c) • E/W Component (Face West) (d)	Discrete-Categorical
Doors-Glass Type (x_{46})	• Clear 1/8, 1/4 inch Air (a) • Clear 1/8, 1/2 inch Air (b) • Clear 1/4, 1/4 inch Air (c) • Clear 1/4, 1/2 inch Air (d)	Discrete-Categorical

In Figure 4.1, the 96-point mixed array and a 96-point Sobol' sequence were combined into a single design using a two-factor Latin hypercube. The purpose of this design approach was to create to a single design that handled both discrete and continuous variables. Column 1 of the Latin hypercube selected rows (points) from the mixed array, and column 2 of the Latin hypercube selected rows (points) from the Sobol' sequence. Hence, each row of the Latin hypercube chooses one point from the mixed array and one point from the Sobol' sequence to create a combination that specifies all 46 factors. For example, the first run of the combined design uses the 65th row of the mixed array and the 37th row of the Sobol' sequence. This specifies all 46 factors for conducting one run of eQUEST. A 96-point Latin hypercube would select each point from the mixed array and each point from the Sobol' sequence exactly once. In this research, 192-point Latin hypercube selects each point twice, specified in Table 4.2 (M: MA, S: Sobol').

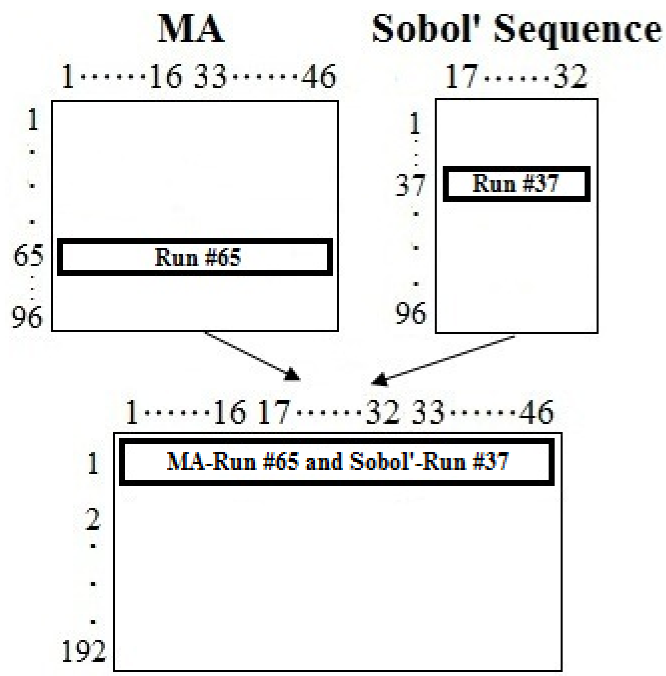


Figure 4.1 Latin Hypercube Design

Table 4.2 Latin Hypercube Design for 192 Runs

Runs	M	S	Runs	M	S	Runs	M	S	Runs	M	S	Runs	M	S
1	65	37	**40**	58	23	**79**	73	2	**118**	82	73	**157**	65	64
2	7	57	**41**	77	86	**80**	89	24	**119**	9	30	**158**	53	66
3	87	73	**42**	21	56	**81**	96	95	**120**	28	59	**159**	84	78
4	69	71	**43**	86	64	**82**	72	11	**121**	44	7	**160**	34	79
5	29	1	**44**	54	90	**83**	42	63	**122**	67	75	**161**	33	86
6	91	62	**45**	16	93	**84**	68	58	**123**	17	60	**162**	41	68
7	57	38	**46**	11	70	**85**	19	40	**124**	40	56	**163**	87	49
8	61	21	**47**	84	41	**86**	70	45	**125**	10	85	**164**	2	82
9	51	44	**48**	44	49	**87**	47	6	**126**	92	6	**165**	56	90
10	40	33	**49**	35	43	**88**	74	76	**127**	95	40	**166**	90	93
11	5	10	**50**	76	36	**89**	43	52	**128**	96	70	**167**	46	45
12	10	69	**51**	48	7	**90**	28	4	**129**	51	87	**168**	60	43
13	71	5	**52**	92	77	**91**	22	74	**130**	12	65	**169**	22	84
14	75	89	**53**	41	16	**92**	25	35	**131**	8	83	**170**	66	77
15	6	30	**54**	2	61	**93**	8	27	**132**	6	89	**171**	61	32
16	64	84	**55**	82	65	**94**	9	17	**133**	71	22	**172**	7	63
17	78	34	**56**	95	42	**95**	49	81	**134**	36	17	**173**	93	62
18	50	92	**57**	24	66	**96**	83	15	**135**	75	20	**174**	68	95
19	32	39	**58**	67	48	**97**	49	13	**136**	64	50	**175**	91	33
20	81	12	**59**	15	85	**98**	55	35	**137**	43	3	**176**	24	36
21	38	60	**60**	79	59	**99**	94	54	**138**	52	52	**177**	85	91
22	27	55	**61**	31	8	**100**	27	11	**139**	79	48	**178**	86	25
23	23	18	**62**	62	53	**101**	69	72	**140**	73	53	**179**	77	18

Table 4.2 – *Continued*

24	17	78	**63**	39	14	**102**	48	51	**141**	29	12	**180**	1	24
25	26	79	**64**	80	32	**103**	47	39	**142**	42	67	**181**	23	74
26	45	88	**65**	33	91	**104**	35	76	**143**	25	21	**182**	45	94
27	63	47	**66**	1	94	**105**	39	23	**144**	30	15	**183**	72	31
28	66	54	**67**	36	51	**106**	54	16	**145**	57	14	**184**	21	37
29	13	46	**68**	46	3	**107**	59	46	**146**	70	71	**185**	50	57
30	88	96	**69**	12	67	**108**	58	61	**147**	13	47	**186**	32	42
31	59	9	**70**	4	20	**109**	83	34	**148**	78	38	**187**	11	69
32	52	26	**71**	56	28	**110**	16	81	**149**	20	4	**188**	37	80
33	30	82	**72**	85	87	**111**	38	44	**150**	19	26	**189**	74	88
34	60	50	**73**	20	31	**112**	4	10	**151**	63	19	**190**	89	5
35	55	13	**74**	94	72	**113**	26	96	**152**	15	28	**191**	31	58
36	3	68	**75**	18	75	**114**	14	1	**153**	18	55	**192**	88	41
37	90	25	**76**	14	22	**115**	76	92	**154**	62	27			
38	93	19	**77**	37	83	**116**	3	2	**155**	5	29			
39	53	29	**78**	34	80	**117**	81	9	**156**	80	8			

4.2 Fitting Treed Regression and TreeMARS

In this dissertation, four models, treed regression (TreeReg), treed multivariate adaptive regression splines (TreeMARS), categorical TreeReg (CATreeReg) and categorical TreeMARS (CATreeMARS), are described. For regression modeling, there are two stepwise regression methods that were employed, Akaike's Information Criterion (AIC) [78-79] and *p*-values. This research employed the R package "step" to choose the predictive variables by an automatic procedure for model selection using minimum AIC, and the mean model is the initial input for variable selection for each performance metric. SAS utilizes *p*-values to yield different variable sets.

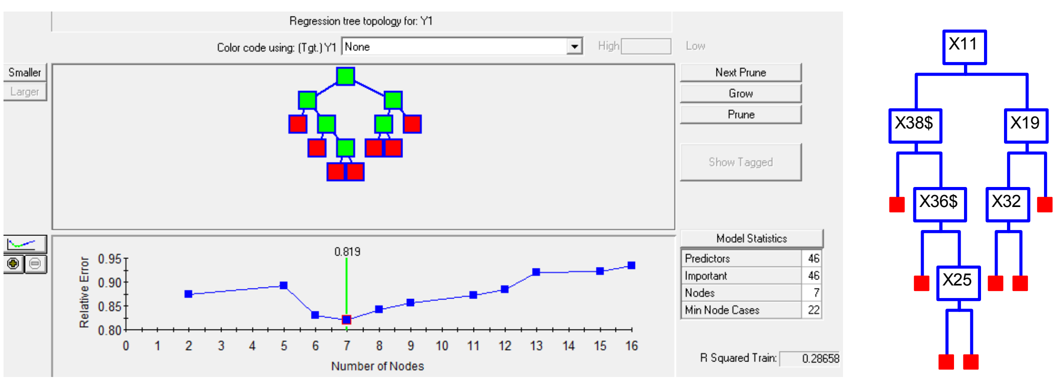
4.2.1 Tree Models

The tree model can be written as

$$\hat{g}_{Tree}(\mathbf{x}) = \sum_{j=1}^{J} \hat{f}(\mathbf{x}) \cdot I\{\mathbf{x} \in R_j\}, \tag{4.2}$$

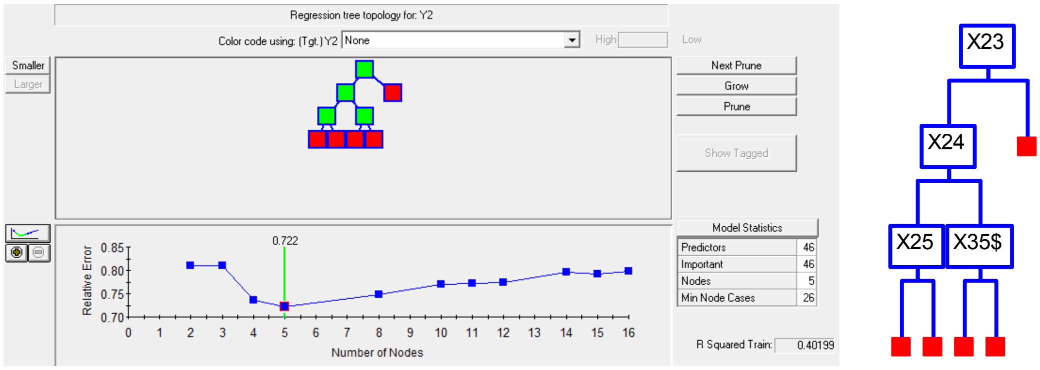
where the regression function $\hat{f}(\mathbf{x})$ is constant, I is an indicator function, $\mathbf{x}$ is the vector of predictor variables, R represents disjoint regions, and J denotes the number of terminal nodes. There are two useful software tools, Salford Systems "CART" and the R package "tree." These two tools include a k-fold cross-validation (CV) technique. While the use of CV with a designed experiment alters the designed structure, the CV approach has been found useful by [80], which used 10-fold CV with a designed experiment. This dissertation also used 10-fold CV. Using only the 192 runs from the second design, CART using CV gave the following message for all four responses: “The optimal tree has no splits.” To remedy this, the 192 design was combined with the 216 initial runs. However, the four factor variables are different between 216 runs and 192 runs. In the 216 runs, there are the four variables Windows-Frame Width, Window Height, Windows-Sill Height and Distance from Window-Overhangs. In the 192 runs, these were replaced by the four variables %Window-North, %Window-South, %Window-East and %Window-West. Since the 216 runs also specified the percent glass for the four sides of the building, the 216 runs and 192 runs could be combined together. In the software CART, the minimum node sizes are 35.

CART Regression tree models with all variables in the tree for four responses with 408 runs are shown in Figures 4.2-4.5, where response 1 is annual source energy, response 2 is HVAC total energy, response 3 is annual total utility cost, and response 4 is life cycle cost. In CART, when only categorical variables are considered in the tree (CATree), the regression tree models show fewer splits in Figures 4.6-4.9.



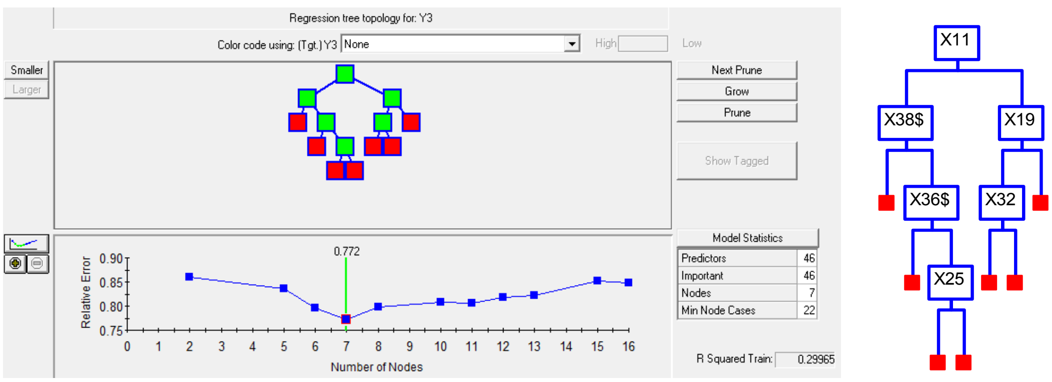
(a) (b)

Figure 4.2 CART. (a) Regression Tree Model for Response 1, (b) Tree Splits



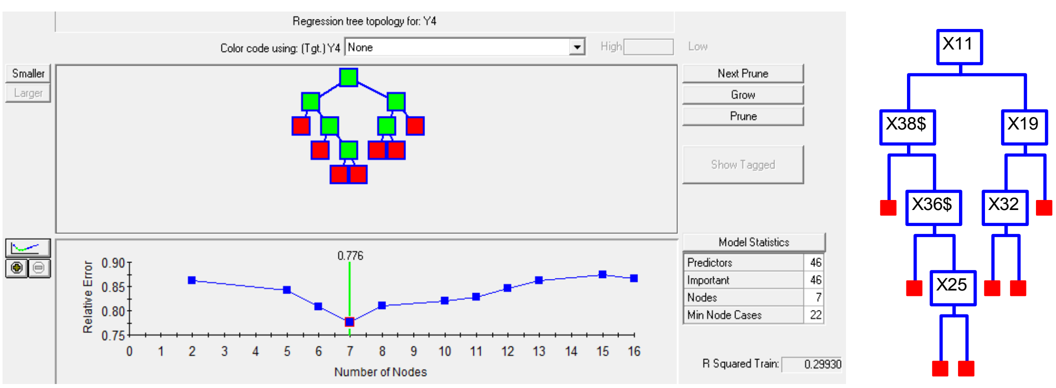
(a) (b)

Figure 4.3 CART. (a) Regression Tree Model for Response 2, (b) Tree Splits



(a) (b)

Figure 4.4 CART. (a) Regression Tree Model for Response 3, (b) Tree Splits



(a) (b)

Figure 4.5 CART. (a) Regression Tree Model for Response 4, (b) Tree Splits

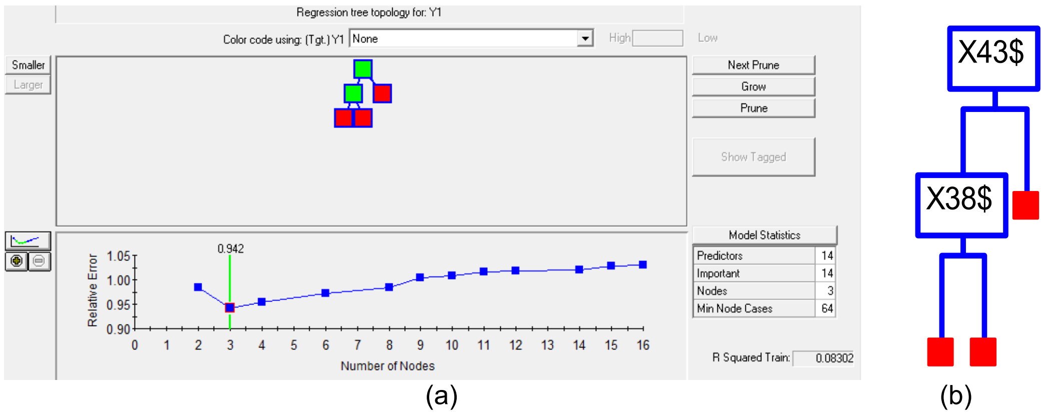


Figure 4.6 CART. (a) Regression CATree Model for Response 1, (b) CATree Splits

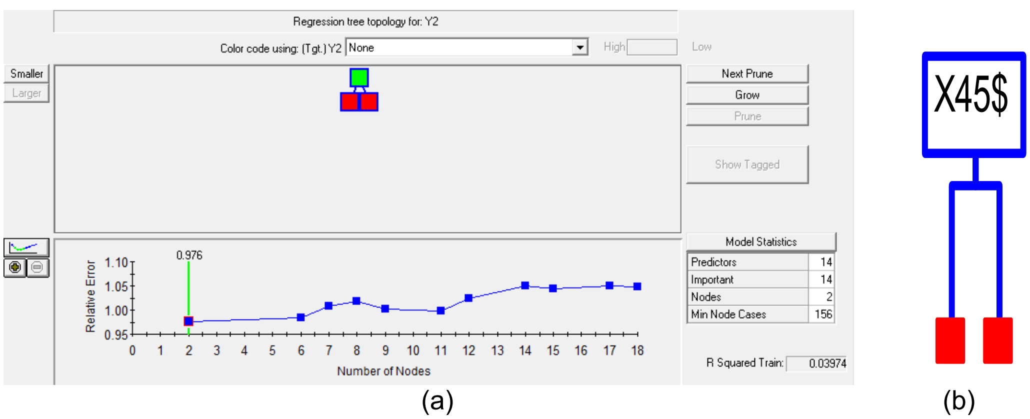


Figure 4.7 CART. (a) Regression CATree Model for Response 2, (b) CATree Splits

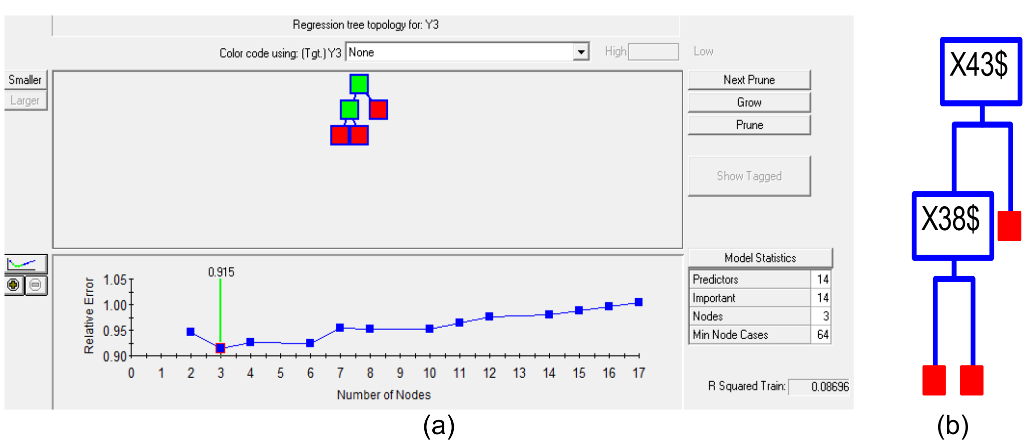
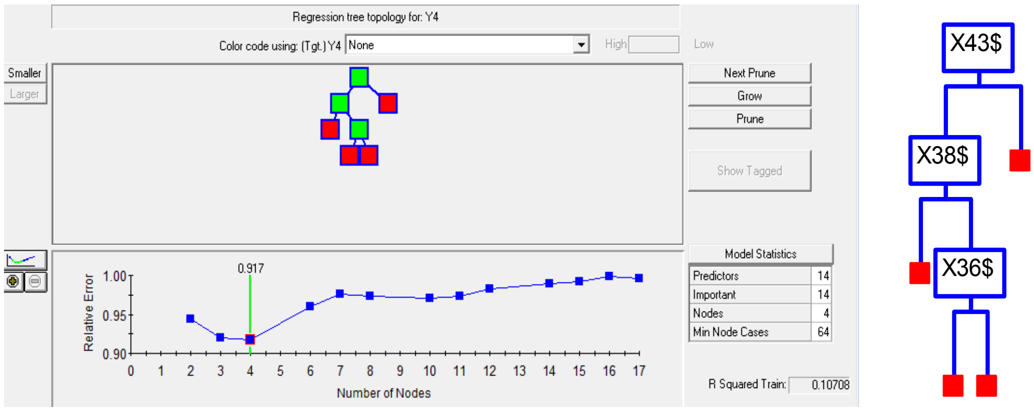


Figure 4.8 CART. (a) Regression CATree Model for Response 3, (b) CATree Splits



(a) (b)

Figure 4.9 CART. (a) Regression CATree Model for Response 4, (b) CATree Splits

Using the R package "tree" [33], the minimum number of observations for each child node (i.e. mincut) is 35. When all variables are considered in the tree model, there are 8 terminal nodes in response 1, there are 7 terminal nodes in response 2, there are 8 terminal nodes in response 3, and there are 8 terminal nodes in response 4. These regression tree models are shown in Figure 4.10. For CATree models, there are 8 terminal nodes in response 1, there are 5 terminal nodes in response 2, there are 8 terminal nodes in response 3, and there are 8 terminal nodes in response 4. These regression CATrees models are shown in Figure 4.11. Obviously, these two figures using R have some similar splits like the previous CART figures. The results of Tables 4.3-4.6 (TN: Terminal Node) are arranged from Figure 4.10 (a), (b), (c) and (d), the results of Tables 4.7-4.10 (TN: Terminal Node) are arranged from Figure 4.11 (a), (b), (c) and (d), and responses 1, 2 and 3 had the same tree splits and observations when mincut is 35.

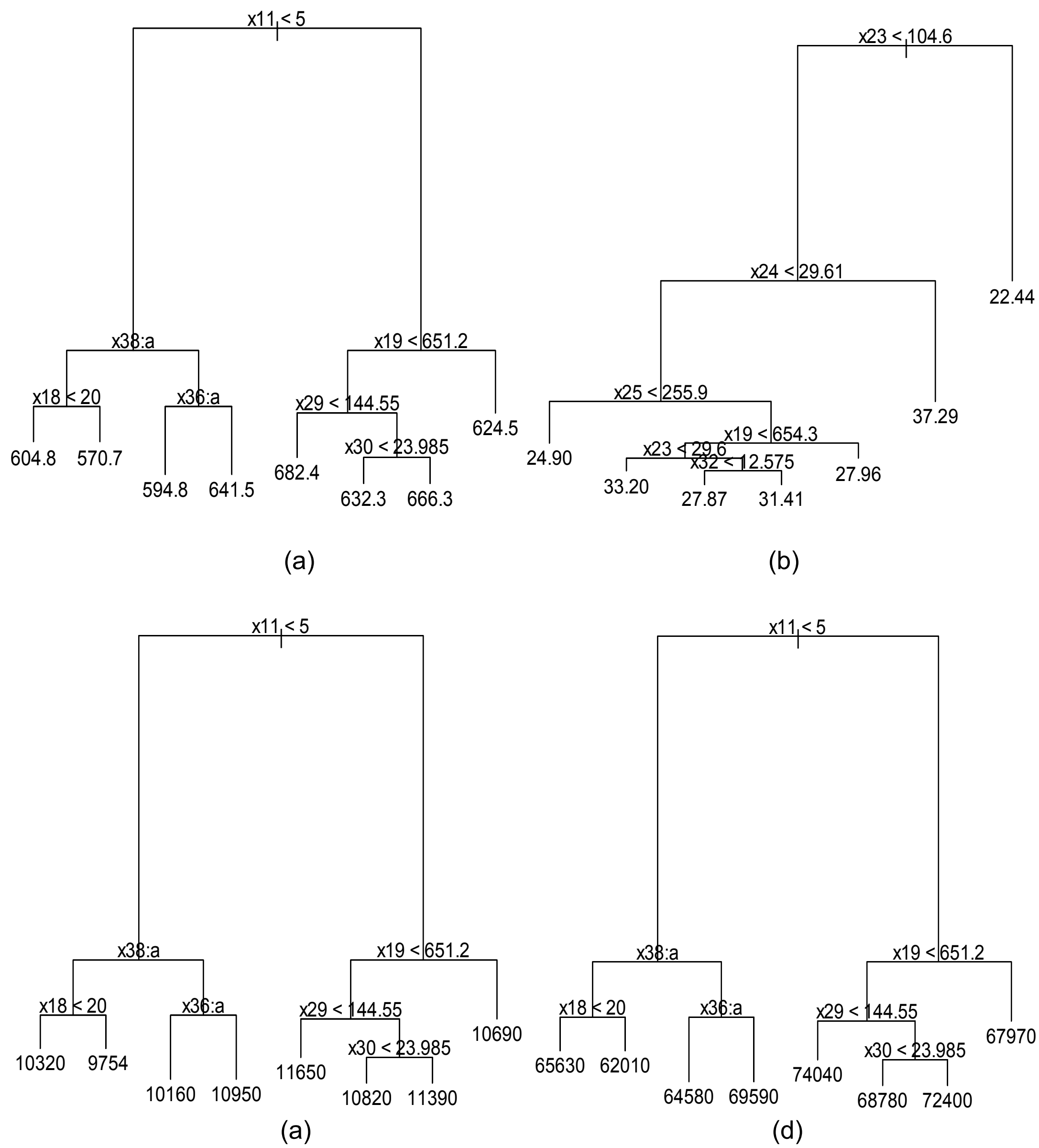


Figure 4.10 Regression Tree Models. (a) Response 1, (b) Response 2, (c) Response 3, (d) Response 4

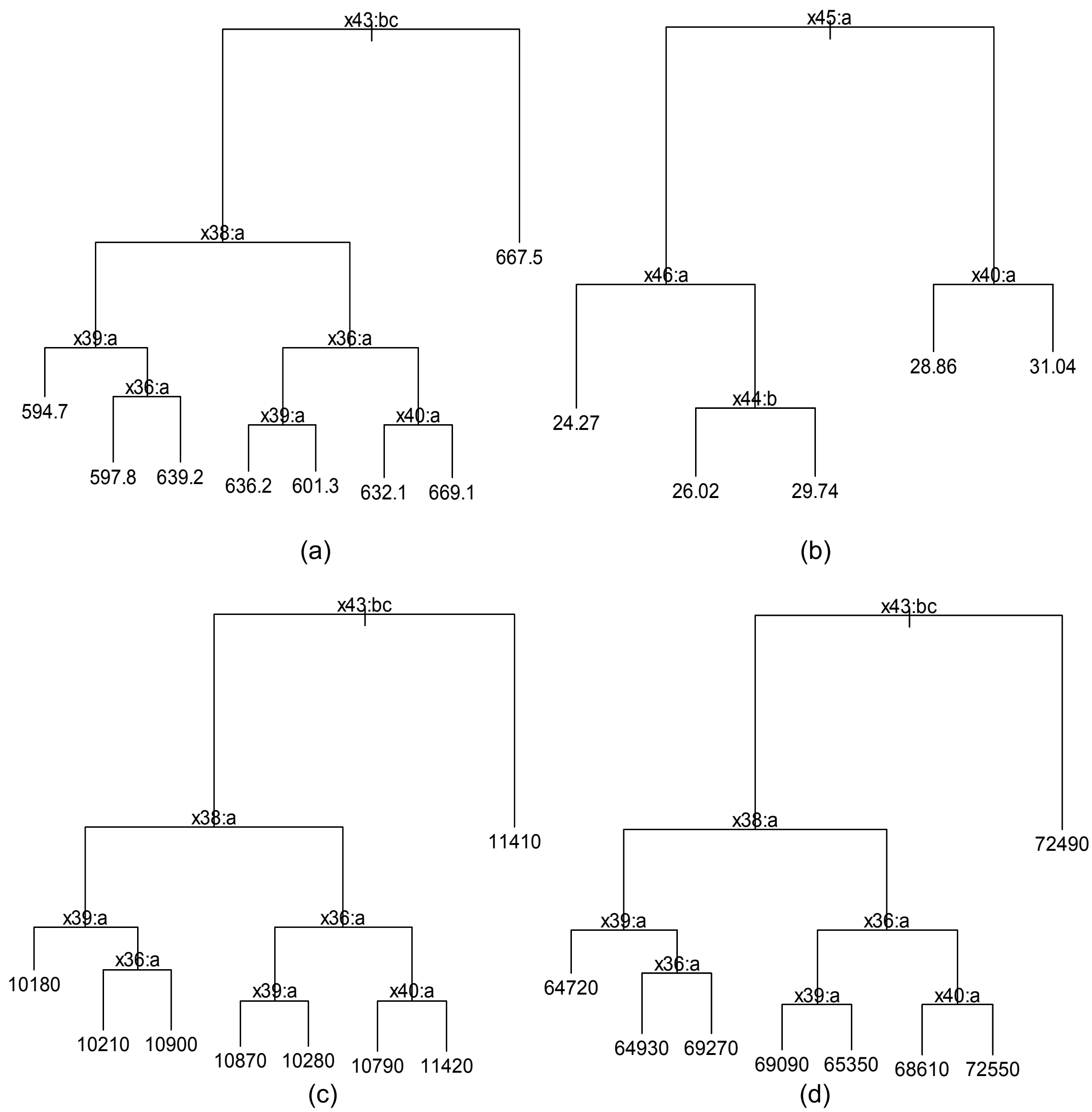


Figure 4.11 Regression CATree Models. (a) Response 1, (b) Response 2, (c) Response 3, (d) Response 4

Table 4.3 Tree Model for Response 1

TN	x_{11}	x_{38}	x_{19}	x_{18}	x_{36}	x_{29}	x_{30}	Mean	Observations
1	< 5	a		< 20				604.8	54
2	< 5	a		> 20				570.7	54
3	< 5	b			a			594.8	54
4	< 5	b			b			641.5	54
5	> 5		< 651.2			< 144.55		682.4	68

Table 4.3 – *Continued*

6	> 5		< 651.2			> 144.55	< 23.985	632.3	38
7	> 5		< 651.2			> 144.55	> 23.985	666.3	42
8	> 5		> 651.2					624.5	44

Table 4.4 Tree Model for Response 2

TN	x_{23}	x_{24}	x_{25}	x_{19}	x_{32}	Mean	Observations
1	< 104.6	< 29.61	< 255.9			24.90	42
2	< 104.6, < 29.6	< 29.61	> 255.9	< 654.3		33.20	46
3	< 104.6, > 29.6	< 29.61	> 255.9	< 654.3	< 12.575	27.87	38
4	< 104.6, > 29.6	< 29.61	> 255.9	< 654.3	> 12.575	31.41	68
5	< 104.6	< 29.61	> 255.9	> 654.3		27.96	50
6	< 104.6	> 29.61				37.29	56
7	> 104.6					22.44	108

Table 4.5 Tree Model for Response 3

TN	x_{11}	x_{38}	x_{19}	x_{18}	x_{36}	x_{29}	x_{30}	Mean	Observations
1	< 5	a		< 20				10320	54
2	< 5	a		> 20				9754	54
3	< 5	b			a			10160	54
4	< 5	b			b			10950	54
5	> 5		< 651.2			< 144.55		11650	68
6	> 5		< 651.2			> 144.55	< 23.985	10820	38
7	> 5		< 651.2			> 144.55	> 23.985	11390	42
8	> 5		> 651.2					10690	44

Table 4.6 Tree Model for Response 4

TN	x_{11}	x_{38}	x_{19}	x_{18}	x_{36}	x_{29}	x_{30}	Mean	Observations
1	< 5	a		< 20				65630	54
2	< 5	a		> 20				62010	54
3	< 5	b			a			64580	54
4	< 5	b			b			69590	54
5	> 5		< 651.2			< 144.55		74040	68
6	> 5		< 651.2			> 144.55	< 23.985	68780	38
7	> 5		< 651.2			> 144.55	> 23.985	72400	42
8	> 5		> 651.2					67970	44

Table 4.7 CATree Model for Response 1

TN	x_{43}	x_{38}	x_{39}	x_{36}	x_{40}	Mean	Observations
1	b, c	a	a			594.7	86
2	b, c	a	b	a		597.8	40

Table 4.7 – *Continued*

3	b, c	a	b	b		639.2	46
4	b, c	b	a	a		636.2	40
5	b, c	b	b	a		601.3	46
6	b, c	b		b	a	632.1	46
7	b, c	b		b	b	669.1	40
8	a					667.5	64

Table 4.8 CATree Model for Response 2

TN	x_{45}	x_{46}	x_{40}	x_{44}	Mean	Observations
1	a	a			24.27	66
2	a	b, c, d		b	26.02	40
3	a	b, c, d		a, c, d	29.74	50
4	b, c, d		a		28.86	126
5	b, c, d		b		31.04	126

Table 4.9 CATree Model for Response 3

TN	x_{43}	x_{38}	x_{39}	x_{36}	x_{40}	Mean	Observations
1	b, c	a	a			10180	86
2	b, c	a	b	a		10210	40
3	b, c	a	b	b		10900	46
4	b, c	b	a	a		10870	40
5	b, c	b	b	a		10280	46
6	b, c	b		b	a	10790	46
7	b, c	b		b	b	11420	40
8	a					11410	64

Table 4.10 CATree Model for Response 4

TN	x_{43}	x_{38}	x_{39}	x_{36}	x_{40}	Mean	Observations
1	b, c	a	a			64720	86
2	b, c	a	b	a		64930	40
3	b, c	a	b	b		69270	46
4	b, c	b	a	a		69090	40
5	b, c	b	b	a		65350	46
6	b, c	b		b	a	68610	46
7	b, c	b		b	b	72550	40
8	a					72490	64

4.2.2 Treed Regression Models

The tree model is shown in equation (4.2). Instead of averaging response values (constants) at the terminal nodes of the tree, TreeReg uses the linear regression model

$$\hat{f}(\mathbf{x}) = \beta_0 + \sum_{j=1}^{J} \beta_j \mathbf{x} \tag{4.3}$$

to handle the mix of discrete and continuous factor variables, where β is the vector of model coefficients, which are estimated by $\hat{\beta}$, and J denotes the number of terminal nodes. The detailed TreeReg approach is described in Alexander and Grimshaw [51].

4.2.3 TreeMARS Models

When curvilinearity is present in the relationships between the continuous factor variables, multivariate adaptive regression splines (MARS) may be a better alternative to regression. This research employed a TreeMARS modeling developed by Sahu [52] in 2011 (Figure 4.12). The TreeMARS model can be written as

$$\hat{g}_{TreeMARS}(\mathbf{x}) = \sum_{j=1}^{J} \beta_j B_j(\mathbf{x}) \cdot I\{\mathbf{x} \in R_j\}, \tag{4.4}$$

where β is the vector of model coefficients, which are estimated by $\hat{\beta}$, B functions are basis functions, I is an indicator function, $\mathbf{x}$ is the vector of predictor variables, R represents disjoint regions, J denotes the number of terminal nodes, and the MARS model is

$$\hat{g}_{MARS}(\mathbf{x}) = \beta_0 + \sum_{j=1}^{J} \beta_j B_j(\mathbf{x}). \tag{4.5}$$

The MARS model is described in Section 2.3.3.1, and the detailed TreeMARS approach is described in Sahu’s dissertation. This is an extension of TreeReg. Instead of fitting linear regression models at the terminal nodes of the tree, TreeMARS uses MARS. MARS is more flexible than linear regression for modeling continuous relationships, and categorical factors are handled only by tree modeling.

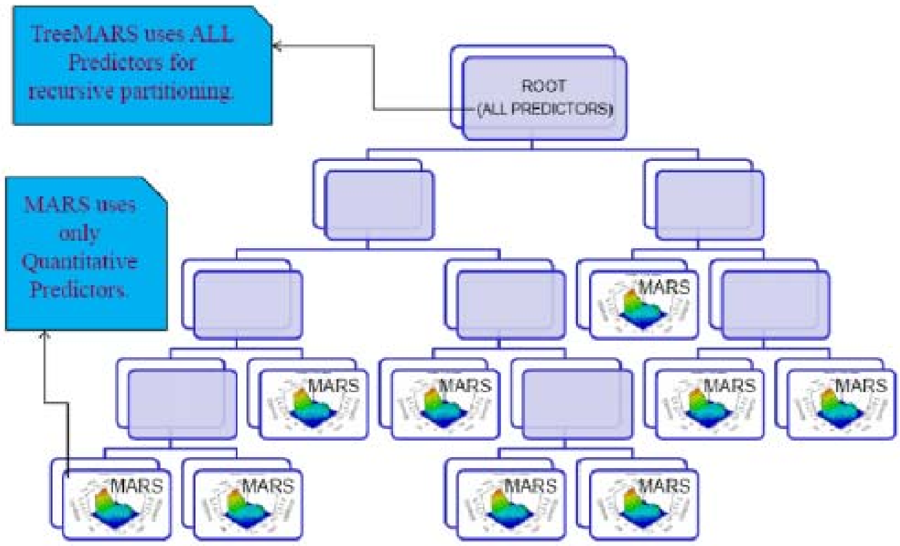


Figure 4.12 Schematic Illustration of TreeMARS Methodology
(Source: Sahu's Dissertation, 2011)

4.3 Handling Multiple Responses

This study has four performance metrics that are potentially correlated. SUR achieves more precise parameter estimates than traditional ordinary least square (OLS) when responses are correlated (Shah et al. [27]). SUR uses the relationships between multiple responses to reduce variances. TreeMARS and TreeReg are statistical linear models, so SUR [27] is applicable. This research fit TreeMARS, TreeReg, CATreeMARS and CATreeReg models for each performance metric to handle mix of discrete and continuous factors, and identify appropriate model forms for each performance metric. The SUR procedure (Shah et al. [27]) is used to generate new parameters, and then re-estimate model parameters using SUR for multiple response variables. For example, Tables 4.11-4.14 (TN: Terminal Node, B: Basis Function) show the parts of coefficients using OLS separately and SUR for four models.

For obtaining important factors, the following 33 factors were identified the TreeMARS models as impacting the green building performance metrics: Ground Floor Construction (x_1), Ground Floor Interior Insulation (x_2), Ground Floor Cap (x_3), Ground Floor Exterior/Cavity Insulation (x_4), Exterior Wall Insulation (x_5), Additional Wall Insulation (x_6), %Window-North (x_7), %Window-South (x_8), %Window-East (x_9), Additional Roof Insulation (x_{11}), Ceiling Batt Insulation (x_{12}), Exterior Roof Insulation (x_{13}), Footprint X (x_{14}), Door Dimension-Width (x_{15}), Door-Frame Width (x_{16}), Design Max Occupant Density-Residential (General Living Space) (x_{17}), Design Ventilation-Residential (General Living Space) (x_{18}), Design Max Occupant Density-Residential (Bedroom) (x_{19}), Design Ventilation-Residential (Bedroom) (x_{20}), Design Max Occupant Density-Residential (Garage) (x_{21}), Design Ventilation-Residential (Garage) (x_{22}), Design Max Occupant Density-Dining Area (x_{23}), Design Ventilation-Dining Area (x_{24}), Design Max Occupant Density-Kitchen and Food Preparation (x_{25}), Design Ventilation-Kitchen and Food Preparation (x_{26}), Design Max Occupant Density-Corridor (x_{27}), Design Ventilation-Corridor (x_{28}), Design Max Occupant Density-Laundry (x_{29}), Design Ventilation-Laundry (x_{30}), Design Max Occupant Density-All Others (x_{31}), Design Ventilation-All Others (x_{32}), Exterior Wall Finishes (x_{36}), and Interior Wall Insulation (x_{38}).

In the TreeReg models using R, the following 34 factors were identified as impacting the green building performance metrics: Ground Floor Construction (x_1), Ground Floor Interior Insulation (x_2), Ground Floor Cap (x_3), Ground Floor Exterior/Cavity Insulation (x_4), Exterior Wall Insulation (x_5), Additional Wall Insulation (x_6), %Window-North (x_7), %Window-South (x_8), %Window-East (x_9), %Window-West (x_{10}), Additional Roof Insulation (x_{11}), Ceiling Batt Insulation (x_{12}), Exterior Roof Insulation (x_{13}), Footprint X (x_{14}), Door Dimension-Width (x_{15}), Door-Frame Width (x_{16}), Design Max Occupant Density-Residential (General Living Space)

(x_{17}), Design Ventilation-Residential (General Living Space) (x_{18}), Design Max Occupant Density-Residential (Bedroom) (x_{19}), Design Ventilation-Residential (Bedroom) (x_{20}), Design Max Occupant Density-Residential (Garage) (x_{21}), Design Ventilation-Residential (Garage) (x_{22}), Design Max Occupant Density-Dining Area (x_{23}), Design Ventilation-Dining Area (x_{24}), Design Max Occupant Density-Kitchen and Food Preparation (x_{25}), Design Ventilation-Kitchen and Food Preparation (x_{26}), Design Max Occupant Density-Corridor (x_{27}), Design Ventilation-Corridor (x_{28}), Design Max Occupant Density-Laundry (x_{29}), Design Ventilation-Laundry (x_{30}), Design Max Occupant Density-All Others (x_{31}), Design Ventilation-All Others (x_{32}), Exterior Wall Finishes (x_{36}), and Interior Wall Insulation (x_{38}). The only difference between TreeMARS and TreeReg is that TreeReg has one more important factor, %Window-West (x_{10}).

In the CATreeMARS models, the following 39 factors were identified as impacting the four green building performance metrics: Ground Floor Construction (x_1), Ground Floor Interior Insulation (x_2), Ground Floor Cap (x_3), Ground Floor Exterior/Cavity Insulation (x_4), Exterior Wall Insulation (x_5), Additional Wall Insulation (x_6), %Window-North (x_7), %Window-South (x_8), %Window-East (x_9), Additional Roof Insulation (x_{11}), Ceiling Batt Insulation (x_{12}), Exterior Roof Insulation (x_{13}), Footprint X (x_{14}), Door Dimension-Width (x_{15}), Door-Frame Width (x_{16}), Design Max Occupant Density-Residential (General Living Space) (x_{17}), Design Ventilation-Residential (General Living Space) (x_{18}), Design Max Occupant Density-Residential (Bedroom) (x_{19}), Design Ventilation-Residential (Bedroom) (x_{20}), Design Max Occupant Density-Residential (Garage) (x_{21}), Design Ventilation-Residential (Garage) (x_{22}), Design Max Occupant Density-Dining Area (x_{23}), Design Ventilation-Dining Area (x_{24}), Design Max

Occupant Density-Kitchen and Food Preparation (x_{25}), Design Ventilation-Kitchen and Food Preparation (x_{26}), Design Max Occupant Density-Corridor (x_{27}), Design Ventilation-Corridor (x_{28}), Design Max Occupant Density-Laundry (x_{29}), Design Ventilation-Laundry (x_{30}), Design Max Occupant Density-All Others (x_{31}), Design Ventilation-All Others (x_{32}), Exterior Wall Finishes (x_{36}), Interior Wall Insulation (x_{38}), Exterior Roof Finish (x_{39}), Exterior Roof Color (x_{40}), Ceiling Interior Finishes (x_{43}), Windows-Glass Type (x_{44}), Orientation (x_{45}), and Doors-Glass Type (x_{46}).

In the CATreeReg modes using R, the following 40 factors were identified as impacting the four green building performance metrics: Ground Floor Construction (x_1), Ground Floor Interior Insulation (x_2), Ground Floor Cap (x_3), Ground Floor Exterior/Cavity Insulation (x_4), Exterior Wall Insulation (x_5), Additional Wall Insulation (x_6), %Window-North (x_7), %Window-South (x_8), %Window-East (x_9), %Window-West (x_{10}), Additional Roof Insulation (x_{11}), Ceiling Batt Insulation (x_{12}), Exterior Roof Insulation (x_{13}), Footprint X (x_{14}), Door Dimension-Width (x_{15}), Door-Frame Width (x_{16}), Design Max Occupant Density-Residential (General Living Space) (x_{17}), Design Ventilation-Residential (General Living Space) (x_{18}), Design Max Occupant Density-Residential (Bedroom) (x_{19}), Design Ventilation-Residential (Bedroom) (x_{20}), Design Max Occupant Density-Residential (Garage) (x_{21}), Design Ventilation-Residential (Garage) (x_{22}), Design Max Occupant Density-Dining Area (x_{23}), Design Ventilation-Dining Area (x_{24}), Design Max Occupant Density-Kitchen and Food Preparation (x_{25}), Design Ventilation-Kitchen and Food Preparation (x_{26}), Design Max Occupant Density-Corridor (x_{27}), Design Ventilation-Corridor (x_{28}), Design Max Occupant Density-Laundry (x_{29}), Design

Ventilation-Laundry (x_{30}), Design Max Occupant Density-All Others (x_{31}), Design Ventilation-All Others (x_{32}), Exterior Wall Finishes (x_{36}), Interior Wall Insulation (x_{38}), Exterior Roof Finish (x_{39}), Exterior Roof Color (x_{40}), Ceiling Interior Finishes (x_{43}), Windows-Glass Type (x_{44}), Orientation (x_{45}), and Doors-Glass Type (x_{46}). The only difference between CATreeMARS and CATreeReg is that CATreeReg has one more important factor, %Window-West (x_{10}).

Table 4.11 Part of Coefficients for TreeMARS

TreeMARS	Coefficient-OLS	Coefficient-SUR
Response 1 Part of TN 1 ⋮ Part of TN 8	[1009.362-71.13585 $\times x_2$] ·I{x_{11}<5}·I{x_{38}:a}·I{x_{18}<20} ⋮ [-8.257779 $\times \mathrm{B}(x_{24}-23.28)$] ·I{$x_{11}$>5}·I{$x_{19}$>651.2}	[1038.696-70.75312 $\times x_2$] ·I{x_{11}<5}·I{x_{38}:a}·I{x_{18}<20} ⋮ [-1.609473 $\times \mathrm{B}(x_{24}-23.28)$] ·I{$x_{11}$>5}·I{$x_{19}$>651.2}
Response 2 Part of TN 1 ⋮ Part of TN 7	[34.9367-4.809841 $\times \mathrm{B}(2-x_3)$] ·I{x_{23}<104.6}·I{x_{24}<29.61} ·I{x_{25}<255.9} ⋮ [0.03030643 $\times \mathrm{B}(x_{31}-100)$] ·I{$x_{23}$>104.6}	[34.97099-5.13517 $\times \mathrm{B}(2-x_3)$] ·I{x_{23}<104.6}·I{x_{24}<29.61} ·I{x_{25}<255.9} ⋮ [0.0289356 $\times \mathrm{B}(x_{31}-100)$] ·I{$x_{23}$>104.6}
Response 3 Part of TN 1 ⋮ Part of TN 8	[12988.78+132.3844 $\times \mathrm{B}(x_1-2)$] ·I{x_{11}<5}·I{x_{38}:a}·I{x_{18}<20} ⋮ [199.5124 $\times \mathrm{B}(x_7-20)$] ·I{$x_{11}$>5}·I{$x_{19}$>651.2}	[13036.4+175.8621 $\times \mathrm{B}(x_1-2)$] ·I{x_{11}<5}·I{x_{38}:a}·I{x_{18}<20} ⋮ [175.5161 $\times \mathrm{B}(x_7-20)$] ·I{$x_{11}$>5}·I{$x_{19}$>651.2}
Response 4 Part of TN 1 ⋮ Part of TN 8	[82557.21+838.9951 $\times \mathrm{B}(x_1-2)$] ·I{x_{11}<5}·I{x_{38}:a}·I{x_{18}<20} ⋮ [1266.996 $\times \mathrm{B}(x_7-20)$] ·I{$x_{11}$>5}·I{$x_{19}$>651.2}	[82861.62+1115.851 $\times \mathrm{B}(x_1-2)$] ·I{x_{11}<5}·I{x_{38}:a}·I{x_{18}<20} ⋮ [1114.493 $\times \mathrm{B}(x_7-20)$] ·I{$x_{11}$>5}·I{$x_{19}$>651.2}

Table 4.12 Part of Coefficients for TreeReg

TreeReg	Coefficient-OLS	Coefficient-SUR
Response 1		
Part of TN 1	[229.64984-0.10581 $\times x_{23}$] $\cdot$I{ x_{11} <5}$\cdot$I{ x_{38} :a}$\cdot$I{ x_{18} <20}	[285.7-0.02122 $\times x_{23}$] $\cdot$I{ x_{11} <5}$\cdot$I{ x_{38} :a}$\cdot$I{ x_{18} <20}
⋮	⋮	⋮
Part of TN 8	[-10.0732 $\times x_4$] $\cdot$I{ x_{11} >5}$\cdot$I{ x_{19} >651.2}	[-5.3784 $\times x_4$] $\cdot$I{ x_{11} >5}$\cdot$I{ x_{19} >651.2}
Response 2		
Part of TN 1	[1799.73157-0.79010 $\times x_{32}$] $\cdot$I{ x_{23} <104.6}$\cdot$I{ x_{24} <29.61} $\cdot$I{ x_{25} <255.9}	[1186.64380-0.67960 $\times x_{32}$] $\cdot$I{ x_{23} <104.6}$\cdot$I{ x_{24} <29.61} $\cdot$I{ x_{25} <255.9}
⋮	⋮	⋮
Part of TN 7	[-0.29235 $\times x_{15}$] $\cdot$I{ x_{23} >104.6}	[-0.28404 $\times x_{15}$] $\cdot$I{ x_{23} >104.6}
Response 3		
Part of TN 1	[5862.910-1.354 $\times x_{23}$] $\cdot$I{ x_{11} <5}$\cdot$I{ x_{38} :a}$\cdot$I{ x_{18} <20}	[6303+0.08236 $\times x_{23}$] $\cdot$I{ x_{11} <5}$\cdot$I{ x_{38} :a}$\cdot$I{ x_{18} <20}
⋮	⋮	⋮
Part of TN 8	[-172.889 $\times x_4$] $\cdot$I{ x_{11} >5}$\cdot$I{ x_{19} >651.2}	[-93.03 $\times x_4$] $\cdot$I{ x_{11} >5}$\cdot$I{ x_{19} >651.2}
Response 4		
Part of TN 1	[9640.054+7.322 $\times x_{23}$] $\cdot$I{ x_{11} <5}$\cdot$I{ x_{38} :a}$\cdot$I{ x_{18} <20}	[10888.0397+14.9736 $\times x_{23}$] $\cdot$I{ x_{11} <5}$\cdot$I{ x_{38} :a}$\cdot$I{ x_{18} <20}
⋮	⋮	⋮
Part of TN 8	[-1097.91 $\times x_4$] $\cdot$I{ x_{11} >5}$\cdot$I{ x_{19} >651.2}	[-590.75 $\times x_4$] $\cdot$I{ x_{11} >5}$\cdot$I{ x_{19} >651.2}

Table 4.13 Part of Coefficients for CATreeMARS

CATreeMARS	Coefficient-OLS	Coefficient-SUR
Response 1		
Part of TN 1	[754.5789+30.18362 $\times \mathrm{B}(2-x_2)$] $\cdot$I{ x_{43} :bc}$\cdot$I{ x_{38} :a}$\cdot$I{ x_{39} :a}	[754.9652+27.42108 $\times \mathrm{B}(2-x_2)$] $\cdot$I{ x_{43} :bc}$\cdot$I{ x_{38} :a}$\cdot$I{ x_{39} :a}
⋮	⋮	⋮
Part of TN 8	[1.41281 $\times \mathrm{B}(166.4-x_{29})$] $\cdot$I{ x_{43} :a}	[0.6248472 $\times \mathrm{B}(166.4-x_{29})$] $\cdot$I{ x_{43} :a}

Table 4.13 – *Continued*

Response 2 Part of TN 1 ⋮ Part of TN 5	[42.64101-2.84277 $\times \mathrm{B}(x_1-4)$] ·I{ x_{45} :a}·I{ x_{46} :a} ⋮ [-0.3657441 $\times \mathrm{B}(14.22-x_{32})$] ·I{ x_{45} :bcd}·I{ x_{40} :b}	[41.40222-2.448417 $\times \mathrm{B}(x_1-4)$] ·I{ x_{45} :a}·I{ x_{46} :a} ⋮ [-0.3565829 $\times \mathrm{B}(14.22-x_{32})$] ·I{ x_{45} :bcd}·I{ x_{40} :b}
Response 3 Part of TN 1 ⋮ Part of TN 8	[11858.49+1329.673 $\times \mathrm{B}(x_2-2)$] ·I{ x_{43} :bc}·I{ x_{38} :a}·I{ x_{39} :a} ⋮ [67.50631 $\times \mathrm{B}(x_{21}-655.5)$] ·I{ x_{43} :a}	[12216.81+1053.734 $\times \mathrm{B}(x_2-2)$] ·I{ x_{43} :bc}·I{ x_{38} :a}·I{ x_{39} :a} ⋮ [59.47099 $\times \mathrm{B}(x_{21}-655.5)$] ·I{ x_{43} :a}
Response 4 Part of TN 1 ⋮ Part of TN 8	[75367.26+8440.36 $\times \mathrm{B}(x_2-2)$] ·I{ x_{43} :bc}·I{ x_{38} :a}·I{ x_{39} :a} ⋮ [428.6871 $\times \mathrm{B}(x_{21}-655.5)$] ·I{ x_{43} :a}	[77643.25+6687.274 $\times \mathrm{B}(x_2-2)$] ·I{ x_{43} :bc}·I{ x_{38} :a}·I{ x_{39} :a} ⋮ [377.629 $\times \mathrm{B}(x_{21}-655.5)$] ·I{ x_{43} :a}

Table 4.14 Part of Coefficients for CATreeReg

CATreeReg	**Coefficient-OLS**	**Coefficient-SUR**
Response 1 Part of TN 1 ⋮ Part of TN 8	[768.1223+7.7361 $\times x_{11}$] ·I{ x_{43} :bc}·I{ x_{38} :a}·I{ x_{39} :a} ⋮ [-2.3763 $\times x_{24}$]·I{ x_{43} :a}	[787.4037+6.9235 $\times x_{11}$] ·I{ x_{43} :bc}·I{ x_{38} :a}·I{ x_{39} :a} ⋮ [-1.5277 $\times x_{24}$]·I{ x_{43} :a}
Response 2 Part of TN 1 ⋮ Part of TN 5	[37.63718-0.05336 $\times x_{23}$] ·I{ x_{45} :a}·I{ x_{46} :a} ⋮ [0.07035 $\times x_{14}$] ·I{ x_{45} :bcd}·I{ x_{40} :b}	[36.88695-0.05276 $\times x_{23}$] ·I{ x_{45} :a}·I{ x_{46} :a} ⋮ [0.06031 $\times x_{14}$] ·I{ x_{45} :bcd}·I{ x_{40} :b}
Response 3 Part of TN 1 ⋮ Part of TN 8	[13051.878+128.419 $\times x_{11}$] ·I{ x_{43} :bc}·I{ x_{38} :a}·I{ x_{39} :a} ⋮ [57.867 $\times x_{10}$]·I{ x_{43} :a}	[13320.837+117.083 $\times x_{11}$] ·I{ x_{43} :bc}·I{ x_{38} :a}·I{ x_{39} :a} ⋮ [10.345 $\times x_{10}$]·I{ x_{43} :a}

Table 4.14 – *Continued*

Response 4 Part of TN 1	$[82978.35+815.60\times x_{11}]$ $\cdot$I{x_{43}:bc}$\cdot$I{x_{38}:a}$\cdot$I{x_{39}:a}	$[84726.63+741.92\times x_{11}]$ $\cdot$I{x_{43}:bc}$\cdot$I{x_{38}:a}$\cdot$I{x_{39}:a}
⋮	⋮	⋮
Part of TN 8	$[367.47\times x_{10}]\cdot$I{$x_{43}$:a}	$[62.81\times x_{10}]\cdot$I{$x_{43}$:a}

4.4 Interpreting the Models

The regression models are used to interpret the impact of the identified factors on the four responses. First of all, the larger number for insulation indicates greater resistance to heat. When builders increase the insulation in the buildings, the four responses, annual source energy (response 1), HVAC energy (response 2), annual utility cost (response 3) and life cycle cost (response 4), decrease for practical sense. There are seven variables for insulation, Ground Floor Interior Insulation (x_2), Ground Floor Exterior/Cavity Insulation (x_4), Exterior Wall Insulation (x_5), Additional Wall Insulation (x_6), Additional Roof Insulation (x_{11}), Ceiling Batt Insulation (x_{12}) and Exterior Roof Insulation (x_{13}). In TreeReg (Tables 4.15-4.22), there are three reasonable cases: x_2 that has all negative signs for response 1, and x_6 and x_{12} that have all negative signs for response 2. Other insulation variables which have both negative and positive signs for each response do not make practical sense. In CATreeReg (Tables 4.23-4.30), there are four reasonable cases: x_2 that has all negative signs for responses 1, 3 and 4, x_5 that has all negative signs for responses 1 and 2, x_{11} that has all negative signs for response 2, and x_{12} that has all negative signs for response 2. Other insulation variables which have both positive and negative signs for each response do not make practical sense. Moreover, using OLS and SUR has different signs in CATreeReg. For example, x_4, x_5 and x_{12} in terminal node 6 of responses 3 and 4. The signs of x_4 using SUR, x_5 using OLS and x_{12} using SUR which have negative coefficients make practical sense.

When builders increase the inches or widths in Ground Floor Construction (x_1), Ground Floor Cap (x_3), Footprint X (x_{14}), Door Dimension-Width (x_{15}) and Door-Frame Width (x_{16}), these materials may absorb the heat, then the four responses increase. In TreeReg, both positive and negative signs appear in different terminal nodes for each response, and the signs do not have any reasonable cases. In CATreeReg, there are two reasonable cases: x_{14} and x_{15} that have positive signs for response 2, other variables which have both positive and negative signs for each response do not make practical sense. Moreover, in CATreeReg, x_{15} in terminal node 6 of responses 3 and 4 using OLS and SUR has different signs. The sign of x_{15} using SUR which has a positive coefficient makes practical sense. x_{16} in terminal node 7 of response 1 using OLS and SUR has different signs. The sign of x_{16} using SUR which has a positive coefficient makes practical sense.

Many windows which are put on East and West facing will cause higher energy use and cost. For example, people may pay more bills for utilities in summer, so windows on North and South facing may reduce energy use and cost. In TreeReg, when the variable %Window-West (x_{10}) increases, responses 1, 3 and 4 increase. Therefore, x_{10} is the only one reasonable case for responses 1, 3 and 4. Others do not make sense, for example, %Window-West (x_{10}) that has a negative coefficient for response 2, and %Window-North (x_7), %Window-South (x_8) and %Window-East (x_9) that have both positive and negative signs that appear in different terminal nodes for each response. In CATreeReg, there are three reasonable cases: x_7 that has negative signs for response 2, x_8 that has negative signs for responses 1, 3 and 4, and x_{10} that has positive signs for responses 1, 3 and 4. x_9 has both positive and negative signs for each response that do not make sense. Moreover, in CATreeReg, x_7 in terminal node 6 of responses 3 and 4 using OLS and SUR has different signs. The sign of x_7 using OLS which has a negative coefficient makes practical sense.

When builders increase occupant density in the buildings, the energy use and cost will increase. In TreeReg, there are two reasonable cases for occupant density: Occupant Density-Dining Area (x_{23}) and Design Max Occupant Density-All Others (x_{31}) have all positive signs for response 2, other design max occupant density variables have both positive and negative signs for each response that do not make practical sense. However, when builders increase ventilation in the buildings, the energy use and cost will decrease. For example, people do not turn on the air conditioner systems. There are three reasonable cases for ventilation, Design Ventilation-Residential (Garage) (x_{22}) that has all negative signs for response 1, and Design Ventilation-Residential (Bedroom) (x_{20}) and Design Ventilation-All Others (x_{32}) that have all negative signs for response 2. Other ventilation variables that have both positive and negative signs for each response do not make practical sense. In TreeReg, Design Ventilation-Laundry (x_{30}) that has different signs in terminal node 1 for response 1 using OLS and SUR. The sign of x_{30} using SUR which has a negative coefficient makes practical sense. In CATreeReg, there are seven reasonable cases: Design Max Occupant Density-Residential (Garage) (x_{21}), Design Max Occupant Density-Kitchen and Food Preparation (x_{25}), Design Max Occupant Density-Corridor (x_{27}) and Design Max Occupant Density-Laundry (x_{29}) that have positive signs for response 2. Design Ventilation-Residential (Garage) (x_{22}) and Design Ventilation-Laundry (x_{30}) that have negative signs for response 2, and Design Ventilation-All Others (x_{32}) that has negative signs for response 1. In CATreeReg, some variables, x_{17} , x_{20} , x_{21} , x_{23} , x_{28} and x_{29} , have different signs using OLS and SUR. The signs of x_{17} using SUR, and x_{21} , x_{23} , x_{29}

using OLS which have positive coefficients make practical sense. The signs of x_{20} and x_{28} using OLS which have negative coefficients make practical sense.

As a whole, there are many reasonable relationships between predictor and response 2, and CATreeReg provided more beneficial relationships than TreeReg.

Table 4.15 Sign of TreeReg for Response 1-OLS

TN	Sign	Variable
1	Positive	$x_1, x_3, x_5, x_7, x_9, x_{14}, x_{17}, x_{21}, x_{27}, x_{29}, x_{30}, x_{31}, x_{32}$
	Negative	$x_2, x_4, x_6, x_{11}, x_{12}, x_{13}, x_{16}, x_{20}, x_{23}, x_{24}, x_{25}, x_{26}, x_{28}$
2	Positive	$x_7, x_9, x_{12}, x_{13}, x_{16}, x_{17}, x_{20}, x_{23}, x_{25}, x_{28}, x_{29}, x_{32}$
	Negative	$x_4, x_5, x_{11}, x_{14}, x_{19}, x_{21}, x_{22}, x_{24}, x_{26}, x_{27}, x_{30}, x_{31}$
3	Positive	$x_3, x_7, x_9, x_{11}, x_{12}, x_{14}, x_{15}, x_{20}, x_{21}, x_{26}$
	Negative	$x_1, x_2, x_4, x_5, x_6, x_{13}, x_{17}, x_{18}, x_{19}, x_{22}, x_{24}, x_{28}, x_{29}, x_{30}, x_{31}, x_{32}$
4	Positive	$x_1, x_3, x_4, x_6, x_{18}, x_{19}, x_{21}, x_{23}, x_{25}, x_{26}, x_{27}, x_{28}, x_{29}$
	Negative	$x_2, x_5, x_{11}, x_{12}, x_{14}, x_{15}, x_{16}, x_{22}, x_{24}, x_{30}, x_{31}$
5	Positive	$x_6, x_{11}, x_{18}, x_{25}, x_{26}$
	Negative	$x_4, x_{12}, x_{19}, x_{20}, x_{24}, x_{32}$
6	Positive	x_6, x_8, x_{10}, x_{23}
	Negative	x_7, x_9, x_{14}
7	Positive	x_6, x_{11}
	Negative	$x_1, x_2, x_4, x_{14}, x_{23}$
8	Positive	x_1, x_7, x_{29}
	Negative	$x_2, x_3, x_4, x_{19}, x_{24}$

Table 4.16 Sign of TreeReg for Response 1-SUR

TN	Sign	Variable
1	Positive	$x_1, x_3, x_5, x_7, x_9, x_{14}, x_{17}, x_{21}, x_{27}, x_{29}, x_{31}, x_{32}$
	Negative	$x_2, x_4, x_6, x_{11}, x_{12}, x_{13}, x_{16}, x_{20}, x_{23}, x_{24}, x_{25}, x_{26}, x_{28}, x_{30}$
2	Positive	$x_7, x_9, x_{12}, x_{13}, x_{16}, x_{17}, x_{20}, x_{23}, x_{25}, x_{28}, x_{29}, x_{32}$

Table 4.16 – *Continued*

	Negative	x_4, x_5, x_{11}, x_{14}, x_{19}, x_{21}, x_{22}, x_{24}, x_{26}, x_{27}, x_{30}, x_{31}
3	Positive	x_3, x_7, x_9, x_{11}, x_{12}, x_{14}, x_{15}, x_{20}, x_{21}, x_{26}
	Negative	x_1, x_2, x_4, x_5, x_6, x_{13}, x_{17}, x_{18}, x_{19}, x_{22}, x_{24}, x_{28}, x_{29}, x_{30}, x_{31}, x_{32}
4	Positive	x_1, x_3, x_4, x_6, x_{18}, x_{19}, x_{21}, x_{23}, x_{25}, x_{26}, x_{27}, x_{28}, x_{29}
	Negative	x_2, x_5, x_{11}, x_{12}, x_{14}, x_{15}, x_{16}, x_{22}, x_{24}, x_{30}, x_{31}
5	Positive	x_6, x_{11}, x_{18}, x_{25}, x_{26}
	Negative	x_4, x_{12}, x_{19}, x_{20}, x_{24}, x_{32}
6	Positive	x_6, x_8, x_{10}, x_{23}
	Negative	x_7, x_9, x_{14}
7	Positive	x_6, x_{11}
	Negative	x_1, x_2, x_4, x_{14}, x_{23}
8	Positive	x_1, x_7, x_{29}
	Negative	x_2, x_3, x_4, x_{19}, x_{24}

Table 4.17 Sign of TreeReg for Response 2-OLS

TN	Sign	Variable
1	Positive	x_2, x_4, x_5, x_9, x_{10}, x_{13}, x_{14}, x_{16}, x_{18}, x_{21}, x_{23}, x_{24}, x_{26}, x_{31}
	Negative	x_3, x_7, x_8, x_{11}, x_{15}, x_{19}, x_{20}, x_{25}, x_{29}, x_{32}
2	Positive	x_5, x_{14}, x_{23}, x_{26}
	Negative	x_3, x_{11}, x_{18}, x_{21}, x_{24}
3	Positive	x_8
	Negative	x_4, x_9, x_{16}, x_{18}, x_{26}
4	Positive	x_4, x_8, x_{11}, x_{14}, x_{18}, x_{23}, x_{27}, x_{28}
	Negative	x_1, x_{12}, x_{13}, x_{20}, x_{21}
5	Positive	x_{21}, x_{29}, x_{30}
	Negative	x_3
6	Positive	x_3, x_{17}, x_{18}, x_{21}, x_{22}, x_{26}, x_{31}
	Negative	x_4, x_6, x_{10}, x_{11}, x_{14}, x_{16}, x_{20}, x_{29}, x_{30}
7	Positive	x_{16}, x_{24}, x_{25}, x_{28}, x_{30}, x_{31}

Table 4.17 – *Continued*

	Negative	$x_1, x_4, x_6, x_9, x_{13}, x_{14}, x_{15}, x_{17}, x_{18}, x_{27}, x_{29}$

Table 4.18 Sign of TreeReg for Response 2-SUR

TN	Sign	Variable
1	Positive	$x_2, x_4, x_5, x_9, x_{10}, x_{14}, x_{16}, x_{18}, x_{21}, x_{23}, x_{24}, x_{26}, x_{31}$
	Negative	$x_3, x_7, x_8, x_{11}, x_{13}, x_{15}, x_{19}, x_{20}, x_{25}, x_{29}, x_{32}$
2	Positive	$x_5, x_{14}, x_{23}, x_{26}$
	Negative	$x_3, x_{11}, x_{18}, x_{21}, x_{24}$
3	Positive	x_8
	Negative	$x_4, x_9, x_{16}, x_{18}, x_{26}$
4	Positive	$x_4, x_8, x_{11}, x_{14}, x_{18}, x_{23}, x_{27}, x_{28}$
	Negative	$x_1, x_{12}, x_{13}, x_{20}, x_{21}$
5	Positive	x_{21}, x_{29}, x_{30}
	Negative	x_3
6	Positive	$x_3, x_{17}, x_{18}, x_{21}, x_{22}, x_{26}, x_{31}$
	Negative	$x_4, x_6, x_{10}, x_{11}, x_{14}, x_{16}, x_{20}, x_{29}, x_{30}$
7	Positive	$x_{16}, x_{24}, x_{25}, x_{28}, x_{30}, x_{31}$
	Negative	$x_1, x_4, x_6, x_9, x_{13}, x_{14}, x_{15}, x_{17}, x_{18}, x_{27}, x_{29}$

Table 4.19 Sign of TreeReg for Response 3-OLS

TN	Sign	Variable
1	Positive	$x_1, x_3, x_5, x_9, x_{14}, x_{17}, x_{21}, x_{29}, x_{30}, x_{31}, x_{32}$
	Negative	$x_2, x_4, x_6, x_{11}, x_{12}, x_{13}, x_{15}, x_{16}, x_{20}, x_{22}, x_{23}, x_{24}, x_{25}, x_{26}, x_{28}$
2	Positive	$x_3, x_6, x_{11}, x_{13}, x_{14}, x_{17}, x_{20}, x_{23}, x_{25}, x_{28}, x_{29}, x_{31}, x_{32}$
	Negative	$x_1, x_2, x_4, x_5, x_9, x_{19}, x_{21}, x_{22}, x_{24}, x_{26}, x_{27}, x_{30}$
3	Positive	$x_2, x_3, x_7, x_9, x_{11}, x_{12}, x_{14}, x_{15}, x_{20}, x_{21}, x_{22}, x_{26}$
	Negative	$x_1, x_4, x_5, x_{13}, x_{16}, x_{17}, x_{19}, x_{23}, x_{24}, x_{28}, x_{29}, x_{30}, x_{31}, x_{32}$
4	Positive	$x_1, x_2, x_3, x_4, x_7, x_9, x_{11}, x_{12}, x_{15}, x_{18}, x_{19}, x_{20}, x_{22}, x_{25}, x_{28}, x_{29}$
	Negative	$x_5, x_6, x_{14}, x_{16}, x_{17}, x_{21}, x_{24}, x_{26}, x_{30}, x_{31}$

Table 4.19 – *Continued*

5	Positive	$x_3, x_6, x_{11}, x_{16}, x_{25}, x_{26}$
	Negative	$x_4, x_{12}, x_{19}, x_{20}$
6	Positive	x_6, x_8, x_{10}, x_{23}
	Negative	x_7, x_9, x_{14}
7	Positive	x_6, x_{11}
	Negative	$x_1, x_2, x_4, x_{14}, x_{23}$
8	Positive	x_1, x_7, x_{29}
	Negative	$x_2, x_3, x_4, x_{19}, x_{24}$

Table 4.20 Sign of TreeReg for Response 3-SUR

TN	Sign	Variable
1	Positive	$x_1, x_3, x_5, x_9, x_{14}, x_{17}, x_{21}, x_{23}, x_{29}, x_{31}, x_{32}$
	Negative	$x_2, x_4, x_6, x_{11}, x_{12}, x_{13}, x_{15}, x_{16}, x_{20}, x_{22}, x_{24}, x_{25}, x_{26}, x_{28}, x_{30}$
2	Positive	$x_3, x_6, x_{11}, x_{13}, x_{14}, x_{17}, x_{20}, x_{23}, x_{25}, x_{28}, x_{29}, x_{31}, x_{32}$
	Negative	$x_1, x_2, x_4, x_5, x_9, x_{19}, x_{21}, x_{22}, x_{24}, x_{26}, x_{27}, x_{30}$
3	Positive	$x_2, x_3, x_7, x_9, x_{11}, x_{12}, x_{14}, x_{15}, x_{20}, x_{21}, x_{22}, x_{26}$
	Negative	$x_1, x_4, x_5, x_{13}, x_{16}, x_{17}, x_{19}, x_{23}, x_{24}, x_{28}, x_{29}, x_{30}, x_{31}, x_{32}$
4	Positive	$x_1, x_2, x_3, x_4, x_7, x_9, x_{11}, x_{12}, x_{15}, x_{18}, x_{19}, x_{20}, x_{22}, x_{25}, x_{28}, x_{29}$
	Negative	$x_5, x_6, x_{14}, x_{16}, x_{17}, x_{21}, x_{24}, x_{26}, x_{30}, x_{31}$
5	Positive	$x_3, x_6, x_{11}, x_{16}, x_{25}, x_{26}$
	Negative	$x_4, x_{12}, x_{19}, x_{20}$
6	Positive	x_6, x_8, x_{10}, x_{23}
	Negative	x_7, x_9, x_{14}
7	Positive	x_6, x_{11}
	Negative	$x_1, x_2, x_4, x_{14}, x_{23}$
8	Positive	x_1, x_7, x_{29}
	Negative	$x_2, x_3, x_4, x_{19}, x_{24}$

Table 4.21 Sign of TreeReg for Response 4-OLS

TN	Sign	Variable
1	Positive	x_3, x_5, x_7, x_9, x_{14}, x_{16}, x_{17}, x_{19}, x_{21}, x_{23}, x_{29}, x_{30}, x_{31}, x_{32}
	Negative	x_1, x_2, x_6, x_{11}, x_{12}, x_{13}, x_{20}, x_{22}, x_{24}, x_{25}, x_{26}, x_{28}
2	Positive	x_3, x_6, x_{11}, x_{13}, x_{14}, x_{17}, x_{20}, x_{23}, x_{25}, x_{28}, x_{29}, x_{31}, x_{32}
	Negative	x_1, x_2, x_4, x_5, x_9, x_{19}, x_{21}, x_{22}, x_{24}, x_{26}, x_{27}, x_{30}
3	Positive	x_2, x_3, x_7, x_9, x_{11}, x_{12}, x_{14}, x_{15}, x_{20}, x_{21}, x_{22}, x_{26}
	Negative	x_1, x_4, x_5, x_{13}, x_{16}, x_{17}, x_{19}, x_{23}, x_{24}, x_{28}, x_{29}, x_{30}, x_{31}, x_{32}
4	Positive	x_1, x_2, x_3, x_4, x_7, x_9, x_{11}, x_{12}, x_{15}, x_{18}, x_{19}, x_{20}, x_{22}, x_{25}, x_{28}, x_{29}
	Negative	x_5, x_6, x_{14}, x_{16}, x_{17}, x_{21}, x_{24}, x_{26}, x_{30}, x_{31}
5	Positive	x_3, x_6, x_{11}, x_{16}, x_{25}, x_{26}
	Negative	x_4, x_{12}, x_{19}, x_{20}
6	Positive	x_6, x_8, x_{10}, x_{23}
	Negative	x_7, x_9, x_{14}
7	Positive	x_6, x_{11}
	Negative	x_1, x_2, x_4, x_{14}, x_{23}
8	Positive	x_1, x_7, x_{29}
	Negative	x_2, x_3, x_4, x_{19}, x_{24}

Table 4.22 Sign of TreeReg for Response 4-SUR

TN	Sign	Variable
1	Positive	x_1, x_3, x_5, x_7, x_9, x_{14}, x_{16}, x_{17}, x_{21}, x_{23}, x_{29}, x_{31}, x_{32}
	Negative	x_2, x_6, x_{11}, x_{12}, x_{13}, x_{19}, x_{20}, x_{22}, x_{24}, x_{25}, x_{26}, x_{28}, x_{30}
2	Positive	x_3, x_6, x_{11}, x_{13}, x_{14}, x_{17}, x_{20}, x_{23}, x_{25}, x_{28}, x_{29}, x_{31}, x_{32}
	Negative	x_1, x_2, x_4, x_5, x_9, x_{19}, x_{21}, x_{22}, x_{24}, x_{26}, x_{27}, x_{30}
3	Positive	x_2, x_3, x_7, x_9, x_{11}, x_{12}, x_{14}, x_{15}, x_{20}, x_{21}, x_{22}, x_{26}
	Negative	x_1, x_4, x_5, x_{13}, x_{16}, x_{17}, x_{19}, x_{23}, x_{24}, x_{28}, x_{29}, x_{30}, x_{31}, x_{32}
4	Positive	x_1, x_2, x_3, x_4, x_7, x_9, x_{11}, x_{12}, x_{15}, x_{18}, x_{19}, x_{20}, x_{22}, x_{25}, x_{28}, x_{29}
	Negative	x_5, x_6, x_{14}, x_{16}, x_{17}, x_{21}, x_{24}, x_{26}, x_{30}, x_{31}

Table 4.22 – *Continued*

5	Positive	$x_3, x_6, x_{11}, x_{16}, x_{25}, x_{26}$
	Negative	$x_4, x_{12}, x_{19}, x_{20}$
6	Positive	x_6, x_8, x_{10}, x_{23}
	Negative	x_7, x_9, x_{14}
7	Positive	x_6, x_{11}
	Negative	$x_1, x_2, x_4, x_{14}, x_{23}$
8	Positive	x_1, x_7, x_{29}
	Negative	$x_2, x_3, x_4, x_{19}, x_{24}$

Table 4.23 Sign of CATreeReg for Response 1-OLS

TN	Sign	Variable
1	Positive	$x_3, x_6, x_7, x_{10}, x_{11}, x_{19}, x_{24}, x_{29}$
	Negative	$x_2, x_4, x_8, x_{12}, x_{15}, x_{17}, x_{18}, x_{22}, x_{23}, x_{25}, x_{32}$
2	Positive	$x_4, x_{15}, x_{20}, x_{22}, x_{23}, x_{24}, x_{26}, x_{30}$
	Negative	x_7, x_{12}, x_{19}
3	Positive	$x_4, x_{11}, x_{22}, x_{25}, x_{30}$
	Negative	$x_9, x_{16}, x_{17}, x_{18}, x_{27}$
4	Positive	$x_1, x_{11}, x_{14}, x_{23}$
	Negative	x_6, x_{22}
5	Positive	$x_1, x_{22}, x_{24}, x_{27}$
	Negative	$x_8, x_9, x_{19}, x_{25}, x_{30}$
6	Positive	$x_3, x_6, x_{17}, x_{18}, x_{23}, x_{25}, x_{26}, x_{27}$
	Negative	x_5, x_9, x_{14}, x_{29}
7	Positive	$x_1, x_6, x_7, x_{13}, x_{14}, x_{15}, x_{22}, x_{23}, x_{28}, x_{29}, x_{30}$
	Negative	$x_2, x_3, x_{11}, x_{16}, x_{17}, x_{19}, x_{20}, x_{24}, x_{26}, x_{27}, x_{31}, x_{32}$
8	Positive	None
	Negative	x_1, x_4, x_{24}, x_{29}

Table 4.24 Sign of CATreeReg for Response 1-SUR

TN	Sign	Variable
1	Positive	x_3, x_6, x_7, x_{10}, x_{11}, x_{19}, x_{24}, x_{29}
	Negative	x_2, x_4, x_8, x_{12}, x_{15}, x_{17}, x_{18}, x_{22}, x_{23}, x_{25}, x_{32}
2	Positive	x_4, x_{15}, x_{20}, x_{22}, x_{23}, x_{24}, x_{26}, x_{30}
	Negative	x_7, x_{12}, x_{19}
3	Positive	x_{11}, x_{22}, x_{25}, x_{30}
	Negative	x_4, x_9, x_{16}, x_{17}, x_{18}, x_{27}
4	Positive	x_1, x_{11}, x_{14}, x_{23}
	Negative	x_6, x_{22}
5	Positive	x_1, x_{22}, x_{24}, x_{27}
	Negative	x_8, x_9, x_{19}, x_{25}, x_{30}
6	Positive	x_3, x_6, x_{17}, x_{18}, x_{23}, x_{25}, x_{26}, x_{27}
	Negative	x_5, x_9, x_{14}, x_{29}
7	Positive	x_1, x_6, x_7, x_{13}, x_{14}, x_{15}, x_{16}, x_{20}, x_{22}, x_{28}, x_{29}, x_{30}
	Negative	x_2, x_3, x_{11}, x_{17}, x_{19}, x_{23}, x_{24}, x_{26}, x_{27}, x_{31}, x_{32}
8	Positive	None
	Negative	x_1, x_4, x_{24}, x_{29}

Table 4.25 Sign of CATreeReg for Response 2-OLS

TN	Sign	Variable
1	Positive	x_{25}, x_{28}
	Negative	x_1, x_9, x_{22}, x_{23}
2	Positive	x_6, x_{15}, x_{25}, x_{29}, x_{32}
	Negative	x_5, x_{10}, x_{11}, x_{28}
3	Positive	x_1, x_4, x_8, x_{13}, x_{15}, x_{16}, x_{21}, x_{24}, x_{26}, x_{27}
	Negative	x_3, x_5, x_7, x_{19}, x_{22}, x_{28}, x_{30}
4	Positive	x_{13}
	Negative	x_7, x_{23}

Table 4.25 – *Continued*

5	Positive	$x_8, x_{12}, x_{14}, x_{24}, x_{25}, x_{29}$
	Negative	$x_6, x_{16}, x_{19}, x_{23}, x_{31}$

Table 4.26 Sign of CATreeReg for Response 2-SUR

TN	Sign	Variable
1	Positive	x_{25}, x_{28}
	Negative	x_1, x_9, x_{22}, x_{23}
2	Positive	$x_6, x_{15}, x_{25}, x_{29}, x_{32}$
	Negative	$x_5, x_{10}, x_{11}, x_{28}$
3	Positive	$x_1, x_4, x_8, x_{13}, x_{15}, x_{16}, x_{21}, x_{24}, x_{26}, x_{27}$
	Negative	$x_3, x_5, x_7, x_{19}, x_{22}, x_{28}, x_{30}$
4	Positive	x_{13}
	Negative	x_7, x_{23}
5	Positive	$x_8, x_{12}, x_{14}, x_{24}, x_{25}, x_{29}$
	Negative	$x_6, x_{16}, x_{19}, x_{23}, x_{31}$

Table 4.27 Sign of CATreeReg for Response 3-OLS

TN	Sign	Variable
1	Positive	$x_3, x_6, x_7, x_{10}, x_{11}, x_{19}, x_{24}, x_{29}$
	Negative	$x_2, x_4, x_8, x_{12}, x_{15}, x_{17}, x_{18}, x_{22}, x_{23}, x_{25}, x_{32}$
2	Positive	$x_4, x_{15}, x_{16}, x_{20}, x_{23}, x_{24}, x_{26}, x_{30}$
	Negative	$x_7, x_9, x_{19}, x_{27}, x_{32}$
3	Positive	$x_{11}, x_{12}, x_{22}, x_{25}, x_{30}$
	Negative	$x_5, x_9, x_{16}, x_{17}, x_{18}, x_{27}$
4	Positive	$x_1, x_{11}, x_{14}, x_{23}$
	Negative	x_6, x_{22}
5	Positive	$x_1, x_{22}, x_{24}, x_{27}$
	Negative	$x_8, x_9, x_{19}, x_{25}, x_{30}$
6	Positive	$x_4, x_6, x_9, x_{10}, x_{12}, x_{13}, x_{18}, x_{19}, x_{21}, x_{25}, x_{26}, x_{27}, x_{29}, x_{31}, x_{32}$

Table 4.27 – *Continued*

	Negative	$x_1, x_2, x_5, x_7, x_8, x_{14}, x_{15}, x_{16}, x_{17}, x_{20}, x_{22}, x_{24}, x_{28}, x_{30}$
7	Positive	$x_3, x_{13}, x_{15}, x_{16}, x_{29}, x_{30}$
	Negative	$x_2, x_7, x_{11}, x_{18}, x_{19}, x_{21}, x_{24}, x_{27}, x_{31}$
8	Positive	x_{10}
	Negative	$x_1, x_3, x_{12}, x_{24}, x_{29}$

Table 4.28 Sign of CATreeReg for Response 3-SUR

TN	Sign	Variable
1	Positive	$x_3, x_6, x_7, x_{10}, x_{11}, x_{19}, x_{24}, x_{29}$
	Negative	$x_2, x_4, x_8, x_{12}, x_{15}, x_{17}, x_{18}, x_{22}, x_{23}, x_{25}, x_{32}$
2	Positive	$x_4, x_{15}, x_{16}, x_{20}, x_{23}, x_{24}, x_{26}, x_{30}$
	Negative	$x_7, x_9, x_{19}, x_{27}, x_{32}$
3	Positive	$x_{11}, x_{12}, x_{22}, x_{25}, x_{30}$
	Negative	$x_5, x_9, x_{16}, x_{17}, x_{18}, x_{27}$
4	Positive	$x_1, x_{11}, x_{14}, x_{23}$
	Negative	x_6, x_{22}
5	Positive	$x_1, x_{22}, x_{24}, x_{27}$
	Negative	$x_8, x_9, x_{19}, x_{25}, x_{30}$
6	Positive	$x_1, x_5, x_6, x_7, x_9, x_{10}, x_{13}, x_{15}, x_{17}, x_{18}, x_{19}, x_{25}, x_{26}, x_{27}, x_{28}, x_{31}, x_{32}$
	Negative	$x_2, x_4, x_8, x_{12}, x_{14}, x_{16}, x_{20}, x_{21}, x_{22}, x_{24}, x_{29}, x_{30}$
7	Positive	$x_3, x_{13}, x_{15}, x_{16}, x_{29}, x_{30}$
	Negative	$x_2, x_7, x_{11}, x_{18}, x_{19}, x_{21}, x_{24}, x_{27}, x_{31}$
8	Positive	x_{10}
	Negative	$x_1, x_3, x_{12}, x_{24}, x_{29}$

Table 4.29 Sign of CATreeReg for Response 4-OLS

TN	Sign	Variable
1	Positive	$x_3, x_6, x_7, x_{10}, x_{11}, x_{19}, x_{24}, x_{29}$
	Negative	$x_2, x_4, x_8, x_{12}, x_{15}, x_{17}, x_{18}, x_{22}, x_{23}, x_{25}, x_{32}$

Table 4.29 – *Continued*

2	Positive	x_4, x_{15}, x_{16}, x_{20}, x_{23}, x_{24}, x_{26}, x_{30}
	Negative	x_7, x_9, x_{19}, x_{27}, x_{32}
3	Positive	x_4, x_{11}, x_{22}, x_{25}, x_{30}
	Negative	x_9, x_{16}, x_{17}, x_{18}, x_{27}
4	Positive	x_1, x_{11}, x_{14}, x_{23}
	Negative	x_6, x_{22}
5	Positive	x_1, x_{22}, x_{24}, x_{27}
	Negative	x_8, x_9, x_{19}, x_{25}, x_{30}
6	Positive	x_4, x_6, x_9, x_{10}, x_{12}, x_{13}, x_{18}, x_{19}, x_{21}, x_{25}, x_{26}, x_{27}, x_{29}, x_{31}, x_{32}
	Negative	x_1, x_2, x_5, x_7, x_8, x_{14}, x_{15}, x_{16}, x_{17}, x_{20}, x_{22}, x_{24}, x_{28}, x_{30}
7	Positive	x_3, x_{13}, x_{15}, x_{16}, x_{29}, x_{30}
	Negative	x_2, x_7, x_{11}, x_{18}, x_{19}, x_{21}, x_{24}, x_{27}, x_{31}
8	Positive	x_{10}
	Negative	x_1, x_3, x_{12}, x_{24}, x_{29}

Table 4.30 Sign of CATreeReg for Response 4-SUR

TN	Sign	Variable
1	Positive	x_3, x_6, x_7, x_{10}, x_{11}, x_{19}, x_{24}, x_{29}
	Negative	x_2, x_4, x_8, x_{12}, x_{15}, x_{17}, x_{18}, x_{22}, x_{23}, x_{25}, x_{32}
2	Positive	x_4, x_{15}, x_{16}, x_{20}, x_{23}, x_{24}, x_{26}, x_{30}
	Negative	x_7, x_9, x_{19}, x_{27}, x_{32}
3	Positive	x_4, x_{11}, x_{22}, x_{25}, x_{30}
	Negative	x_9, x_{16}, x_{17}, x_{18}, x_{27}
4	Positive	x_1, x_{11}, x_{14}, x_{23}
	Negative	x_6, x_{22}
5	Positive	x_1, x_{22}, x_{24}, x_{27}
	Negative	x_8, x_9, x_{19}, x_{25}, x_{30}
6	Positive	x_1, x_5, x_6, x_7, x_9, x_{10}, x_{13}, x_{15}, x_{17}, x_{18}, x_{19}, x_{25}, x_{26}, x_{27}, x_{28}, x_{31}, x_{32}
	Negative	x_2, x_4, x_8, x_{12}, x_{14}, x_{16}, x_{20}, x_{21}, x_{22}, x_{24}, x_{29}, x_{30}

Table 4.30 – *Continued*

7	Positive	x_3, x_{13}, x_{15}, x_{16}, x_{29}, x_{30}
	Negative	x_2, x_7, x_{11}, x_{18}, x_{19}, x_{21}, x_{24}, x_{27}, x_{31}
8	Positive	x_{10}
	Negative	x_1, x_3, x_{12}, x_{24}, x_{29}

4.5 Comparing Standard Errors

To compute the standard errors, Shah et al. [27] described two equations. First, the variance of the predicted value of the i th response using OLS is

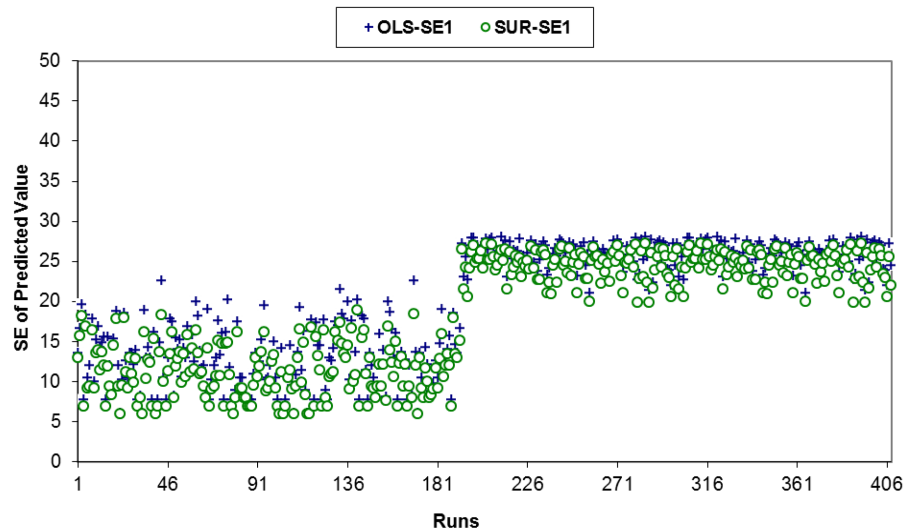
$$\mathrm{Var}[\hat{\mathbf{y}}_i(\mathbf{x})] = \sigma_i^2 \mathbf{f}_i^{\mathrm{T}}(\mathbf{x})(\mathbf{X}_i^{\mathrm{T}}\mathbf{X}_i)^{-1}\mathbf{f}_i(\mathbf{x}),\ i = 1, 2, 3, 4, \tag{4.6}$$

where σ_i^2 is the unknown variance of the random error term. In TreeMARS, the estimator $\hat{\sigma}_i^2$ can be obtained by OLS, the point of interest $\mathbf{f}_i^{\mathrm{T}}(\mathbf{x})$, which is a vector that plugs in the model form at the point $\mathbf{x}$, denotes each row value of basis functions of each performance metric, and $\mathbf{X}_i$ denotes each combination of basis functions for each performance metric. In TreeReg, the estimator $\hat{\sigma}_i^2$ also can be obtained by OLS, $\mathbf{f}_i^{\mathrm{T}}(\mathbf{x})$ denotes each row value of the regression models for each performance metric, and $\mathbf{X}_i$ denotes each combination of regression models for each performance metric.

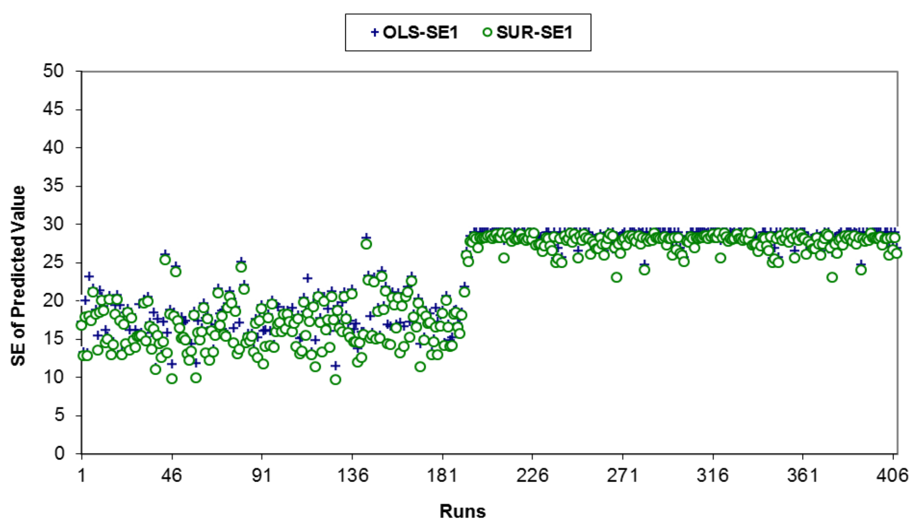
Second, the variance-covariance matrix of predicted response $\hat{\mathbf{y}}(\mathbf{x})$ using SUR is

$$\mathrm{Var}[\hat{\mathbf{y}}(\mathbf{x})] = \mathbf{\Lambda}^{\mathrm{T}}(\mathbf{x})[\mathbf{X}^{\mathrm{T}}(\hat{\mathbf{\Sigma}}^{-1} \otimes \mathbf{I}_n)\mathbf{X}]^{-1}\mathbf{\Lambda}(\mathbf{x}). \tag{4.7}$$

In TreeMARS, $\mathbf{\Lambda}^{\mathrm{T}}(\mathbf{x})$ which is a block-diagonal matrix $\mathrm{diag}(\mathbf{f}_1(\mathbf{x}), \mathbf{f}_2(\mathbf{x}), \mathbf{f}_3(\mathbf{x}), \mathbf{f}_4(\mathbf{x}))$ which denotes each row value of basis functions for four performance metrics, and $\mathbf{X}$ denotes a combination of basis functions for four performance metrics. In TreeReg, $\mathbf{\Lambda}^{\mathrm{T}}(\mathbf{x})$ denotes each row value of the regression models for four performance metrics, and $\mathbf{X}$ denotes a combination of regression models for four performance metrics. Other notation was described in the literature review of SUR. The standard error results for TreeMARS and TreeReg are shown in Figures 4.13-4.16, where it is seen that SUR has consistently smaller standard error values than OLS. The standard error results for CATreeMARS and CATreeReg are shown in Figures 4.17-4.20, and SUR again has smaller standard errors than OLS. This dissertation also compared the standard errors using paired t-tests, and the results are shown in Tables 4.31-4.34. The *p*-values of paired t-tests are all statistically significant, so TreeMARS, TreeReg, CATreeMARS and CATreeReg models with SUR have smaller standard errors.

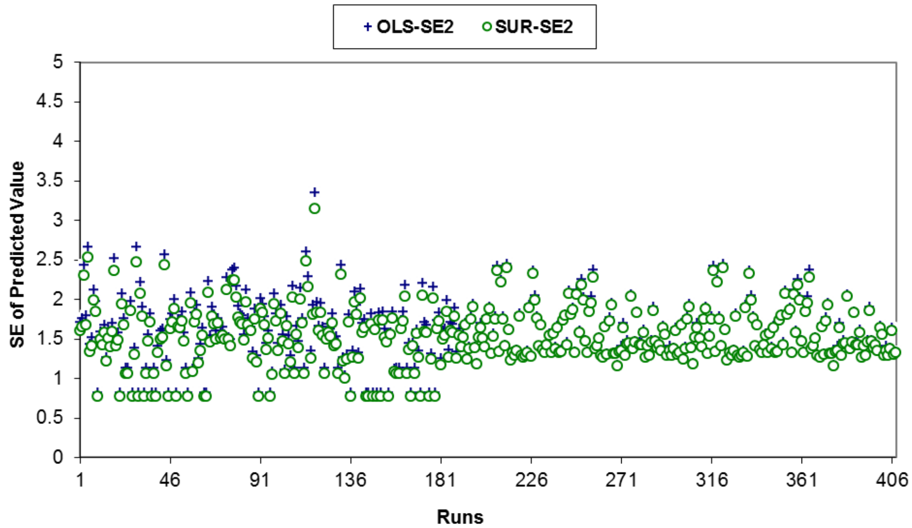


(a)

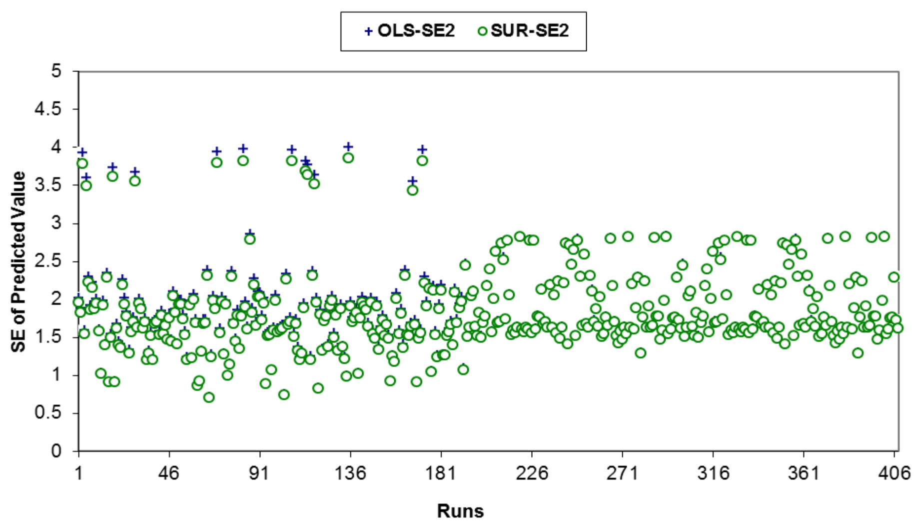


(b)

Figure 4.13 Standard Errors for Response 1 (a) TreeMARS (b) TreeReg

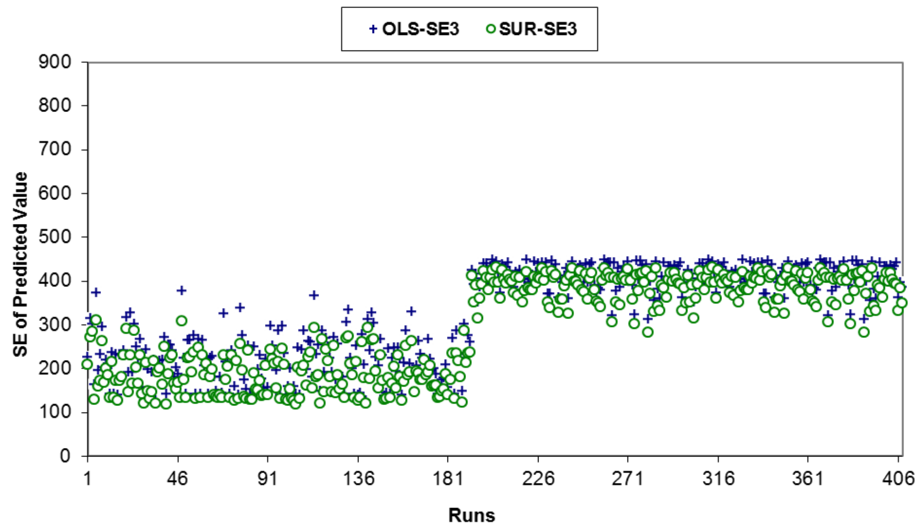


(a)

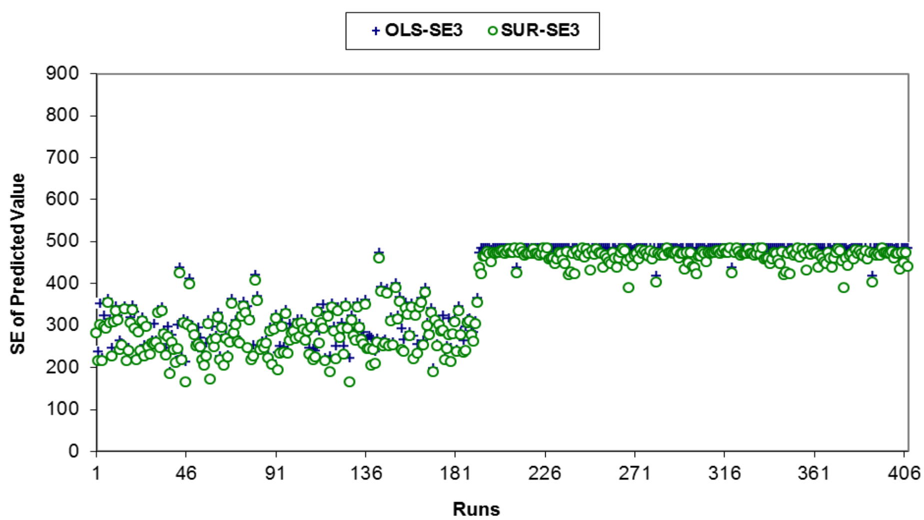


(b)

Figure 4.14 Standard Errors for Response 2 (a) TreeMARS (b) TreeReg

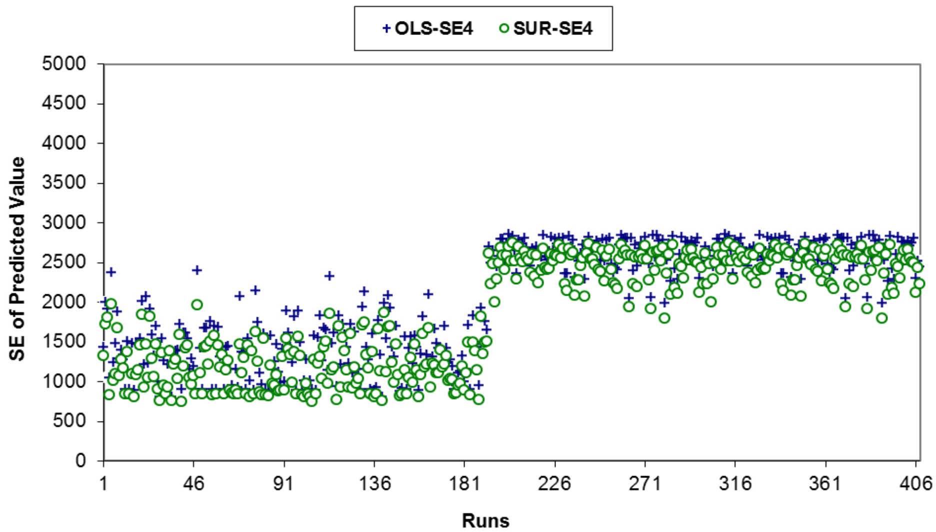


(a)

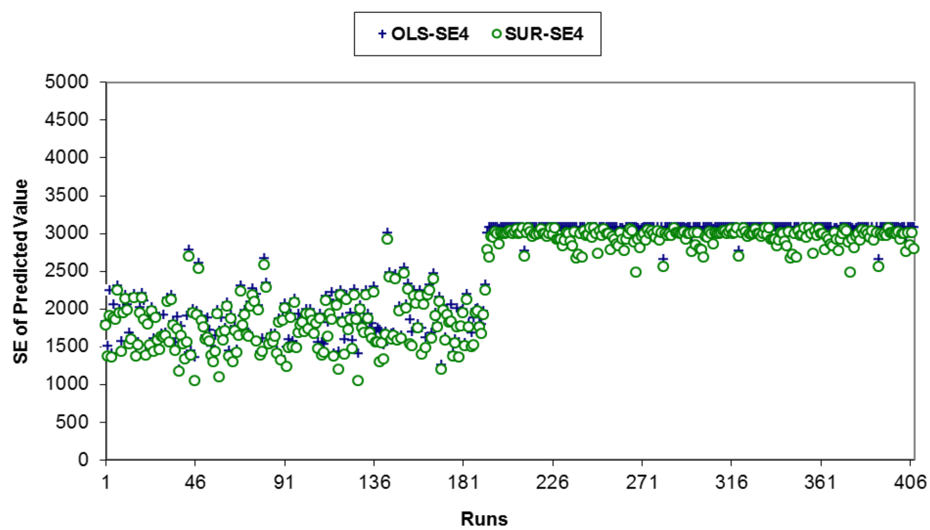


(b)

Figure 4.15 Standard Errors for Response 3 (a) TreeMARS (b) TreeReg

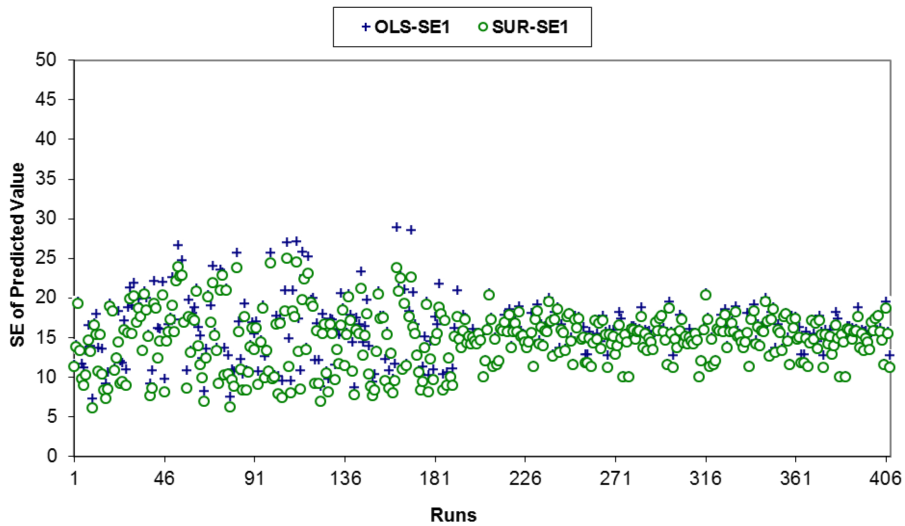


(a)

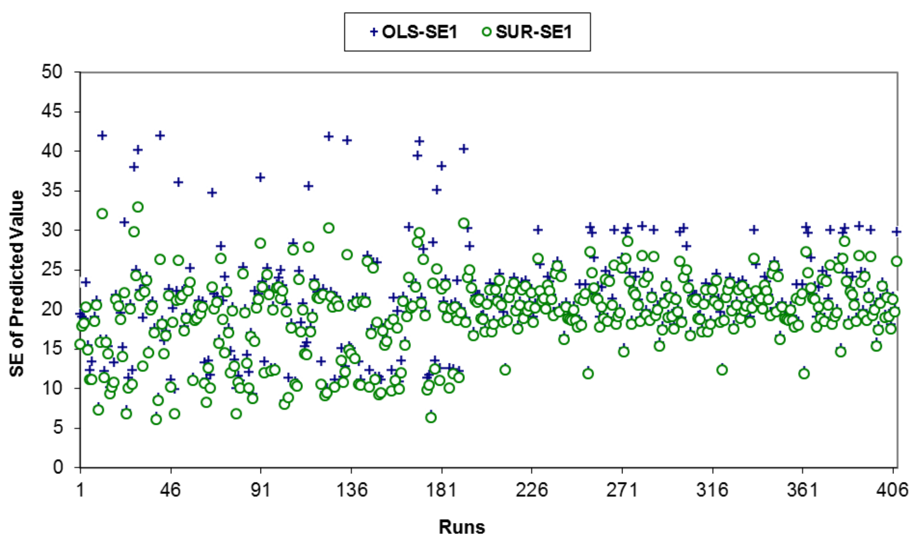


(b)

Figure 4.16 Standard Errors for Response 4 (a) TreeMARS (b) TreeReg

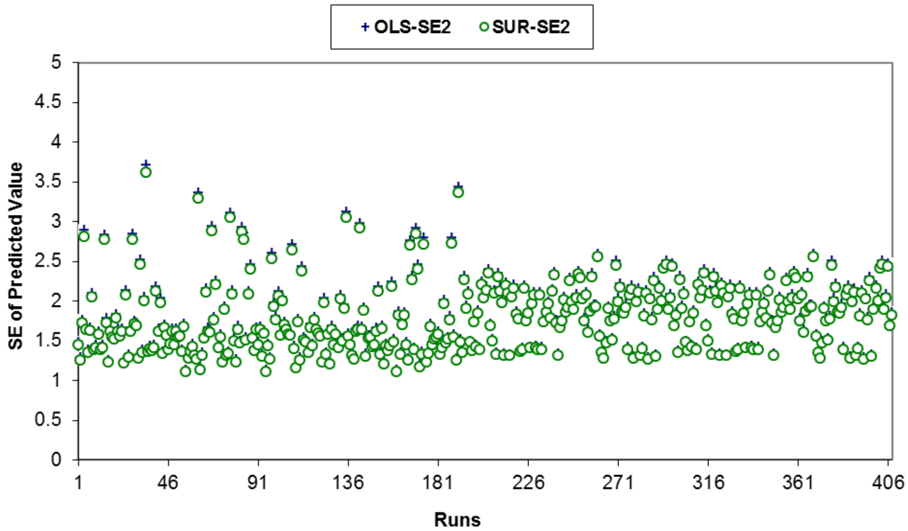


(a)

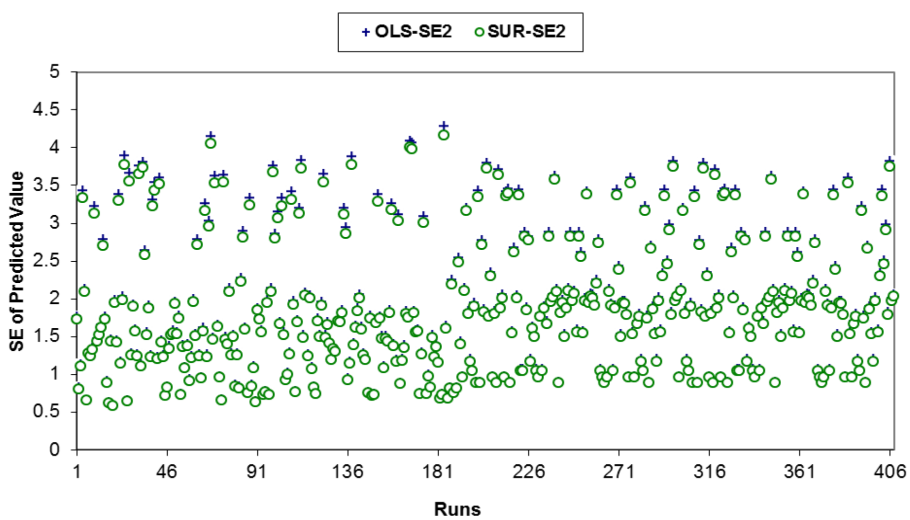


(b)

Figure 4.17 Standard Errors for Response 1 (a) CATreeMARS (b) CATreeReg

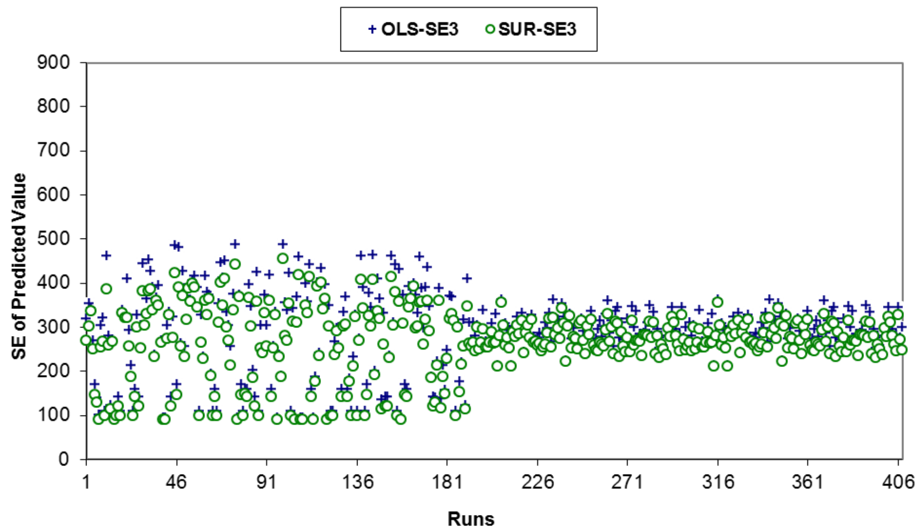


(a)

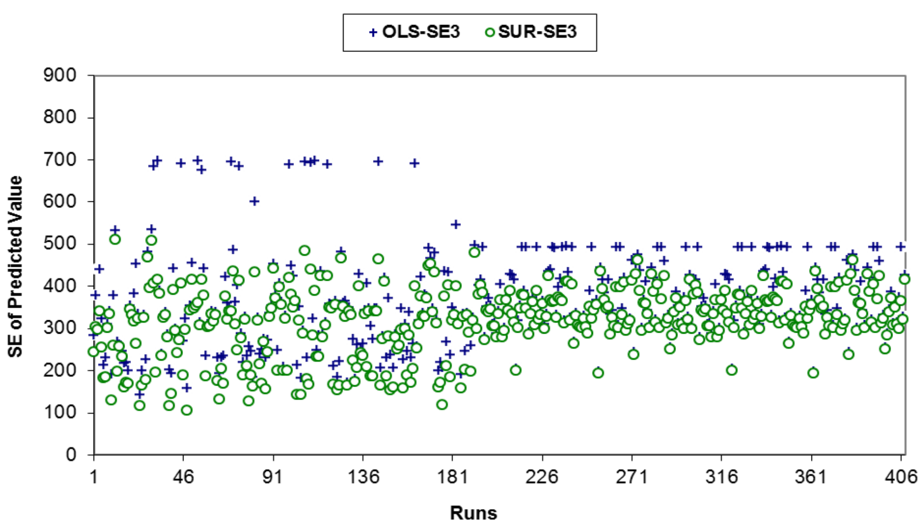


(b)

Figure 4.18 Standard Errors for Response 2 (a) CATreeMARS (b) CATreeReg

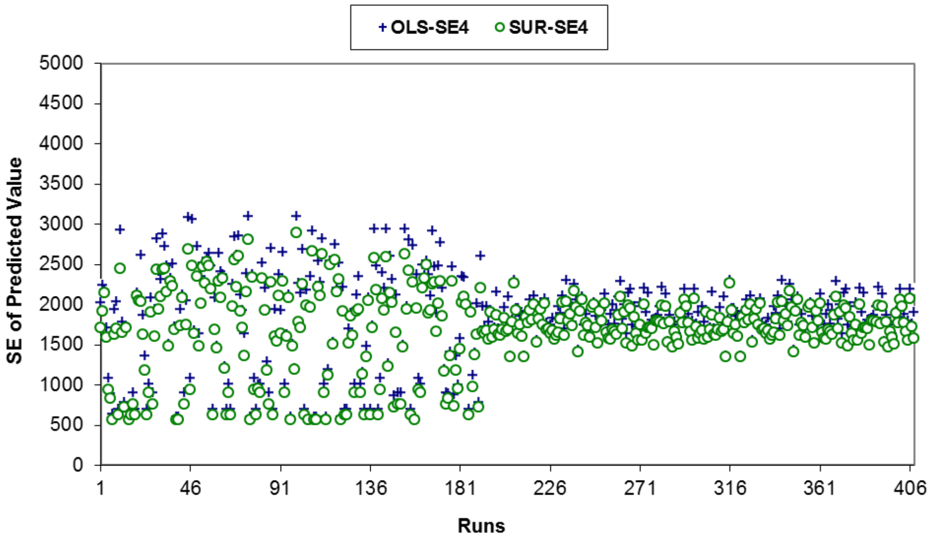


(a)

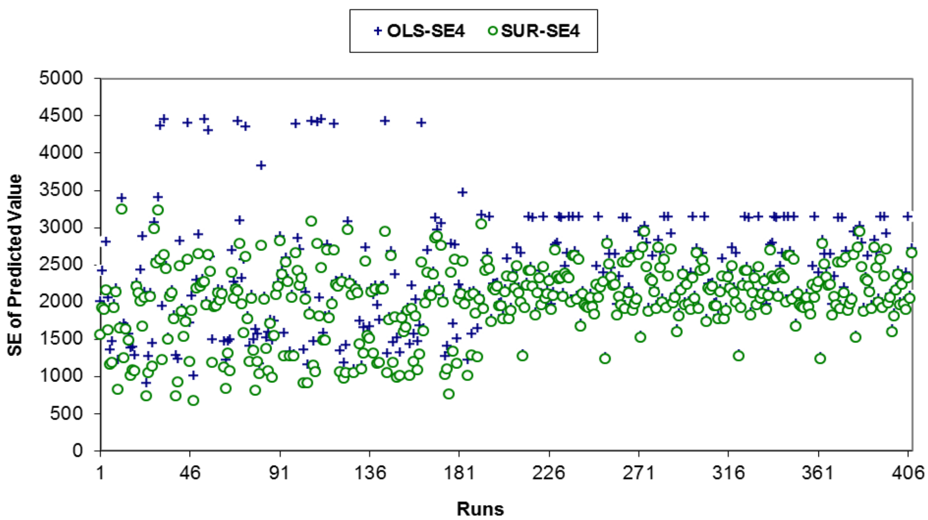


(b)

Figure 4.19 Standard Errors for Response 3 (a) CATreeMARS (b) CATreeReg



(a)



(b)

Figure 4.20 Standard Errors for Response 4 (a) CATreeMARS (b) CATreeReg

Table 4.31 Paired T-Test of Standard Errors for TreeMARS

Response	Mean of Differences	Paired t-test Value	P-Value
1	1.469508	30.4498	< 2.2e-16
2	0.05676378	26.6243	< 2.2e-16
3	28.79566	36.3103	< 2.2e-16
4	183.2271	36.3898	< 2.2e-16

Table 4.32 Paired T-Test of Standard Errors for TreeReg Using R

Response	Mean of Differences	Paired t-test Value	P-Value
1	0.9349627	21.8861	< 2.2e-16
2	0.02833271	17.9946	< 2.2e-16
3	17.49267	23.0846	< 2.2e-16
4	110.9108	23.0508	< 2.2e-16

Table 4.33 Paired T-Test of Standard Errors for CATreeMARS

Response	Mean of Differences	Paired t-test Value	P-Value
1	1.018493	19.7406	< 2.2e-16
2	0.03630752	63.5774	< 2.2e-16
3	24.36132	24.5632	< 2.2e-16
4	155.1313	24.6121	< 2.2e-16

Table 4.34 Paired T-Test of Standard Errors for CATreeReg Using R

Response	Mean of Differences	Paired t-test Value	P-Value
1	1.532993	13.5307	< 2.2e-16
2	0.03834461	34.3509	< 2.2e-16
3	41.65577	14.2308	< 2.2e-16
4	258.3892	13.9235	< 2.2e-16

4.6 Comparing Test Errors

The training models were obtained from the previous 408 runs. To calculate test errors, this research employed an OA 2^{35} design that only has 2 settings and a Sobol' sequence. A 36-point Latin hypercube was used to select each point from the OA and from the Sobol' sequence exactly once, using the same procedure as that described in Section 4.1. These settings and design are shown in Table 4.35 and Table 4.36 (O: OA, and S: Sobol').

Table 4.35 Two Settings for Testing

Variables	**Settings**	**Types**
Ground Floor Construction (x_1)	• 2 inch Concrete • 4 inch Concrete	Discrete-Numerical
Ground Floor Interior Insulation (x_2)	• 1 inch Polystyrene • 1 1/2 inch Polystyrene	Discrete-Numerical

Table 4.35 – *Continued*

Ground Floor Cap (x_3)	• 1.25 inch Lightweight Concrete • 2 inch Lightweight Concrete	Discrete-Numerical
Ground Floor Exterior/Cavity Insulation (x_4)	• 1 inch Polystyrene • 2 inch Polystyrene	Discrete-Numerical
Exterior Wall Insulation (x_5)	• 1 inch Polystyrene • 1 1/2 inch Polystyrene	Discrete-Numerical
Additional Wall Insulation (x_6)	• R-3 Batt • R-7 Batt	Discrete-Numerical
%Window-North (x_7)	• 10 • 25	Discrete-Numerical
%Window-South (x_8)	• 10 • 25	Discrete-Numerical
%Window-East (x_9)	• 10 • 25	Discrete-Numerical
%Window-West (x_{10})	• 10 • 25	Discrete-Numerical
Additional Roof Insulation (x_{11})	• R-7 Batt • R-19 Batt	Discrete-Numerical
Ceiling Batt Insulation (x_{12})	• R-13 Batt • R-19 Batt	Discrete-Numerical
Exterior Roof Insulation (x_{13})	• 1 inch Polystyrene • 1 1/2 inch Polystyrene	Discrete-Numerical
Footprint X (x_{14})	• 100 • 50	Discrete-Numerical
Door Dimension-Width (x_{15})	• 3 • 6	Discrete-Numerical
Door-Frame Width (x_{16})	• 2 • 3	Discrete-Numerical
Design Max Occupant Density-Residential (General Living Space) (x_{17})	Range: 575 to 675	Continuous
Design Ventilation-Residential (General Living Space) (x_{18})	Range: 10 to 30	Continuous
Design Max Occupant Density-Residential (Bedroom) (x_{19})	Range: 575 to 675	Continuous
Design Ventilation-Residential (Bedroom) (x_{20})	Range: 10 to 30	Continuous
Design Max Occupant Density-Residential (Garage) (x_{21})	Range: 575 to 675	Continuous
Design Ventilation-Residential (Garage) (x_{22})	Range: 10 to 30	Continuous
Design Max Occupant Density-Dining Area (x_{23})	Range: 5 to 105	Continuous
Design Ventilation-Dining Area (x_{24})	Range: 10 to 30	Continuous

Table 4.35 – *Continued*

Design Max Occupant Density-Kitchen and Food Preparation (x_{25})	Range: 250 to 350	Continuous
Design Ventilation-Kitchen and Food Preparation (x_{26})	Range: 5 to 25	Continuous
Design Max Occupant Density-Corridor (x_{27})	Range: 100 to 200	Continuous
Design Ventilation-Corridor (x_{28})	Range: 5 to 25	Continuous
Design Max Occupant Density-Laundry (x_{29})	Range: 100 to 200	Continuous
Design Ventilation-Laundry (x_{30})	Range: 15 to 35	Continuous
Design Max Occupant Density-All Others (x_{31})	Range: 100 to 200	Continuous
Design Ventilation-All Others (x_{32})	Range: 5 to 25	Continuous
Wall Construction (x_{33})	• Wood Frame, 2×4, 16 inch o.c. (a) • Wood Frame, 2×4, 24 inch o.c. (b)	Discrete-Categorical
Windows-Glass Category (x_{34})	• Double Clear/Tint (a) • Double Low-e (e2 = 0.1) (b)	Discrete-Categorical
Roof Construction (x_{35})	• Wood Advanced Frame, 24 inch o.c. (a) • Wood Advanced Frame, >24 inch o.c. (b)	Discrete-Categorical
Exterior Wall Finishes (x_{36})	• Brick (a) • Concrete (b)	Discrete-Categorical
Exterior Wall Color (x_{37})	• Light (a) • Dark (b)	Discrete-Categorical
Interior Wall Insulation (x_{38})	• None (a) • 1 inch Polystyrene (b)	Discrete-Categorical
Exterior Roof Finish (x_{39})	• Concrete (a) • Built-up Roof (b)	Discrete-Categorical
Exterior Roof Color (x_{40})	• Light (a) • Dark (b)	Discrete-Categorical
Doors-Construction (x_{41})	• Double Clear/Tint (a) • Double Low-e (e2 = 0.1) (b)	Discrete-Categorical
Pitched Roof (x_{42})	• Without Pitched Roof (a) • With Pitched Roof (b)	Discrete-Categorical
Ceiling Interior Finishes (x_{43})	• Drywall Finish (b) • Plaster Finish (c)	Discrete-Categorical
Windows-Glass Type (x_{44})	• Clear 1/8, 1/4 inch Air (a) • Clear 1/8, 1/2 inch Air (b)	Discrete-Categorical
Orientation (x_{45})	• N/S Component (Face North) (a) • E/W Component (Face East) (c)	Discrete-Categorical
Doors-Glass Type (x_{46})	• Clear 1/8, 1/4 inch Air (a) • Clear 1/8, 1/2 inch Air (b)	Discrete-Categorical

Table 4.36 Latin Hypercube Design for 36 Runs

Runs	O	S	Runs	O	S
1	33	28	**19**	29	19
2	21	26	**20**	4	2
3	26	30	**21**	8	21
4	11	3	**22**	32	7
5	16	17	**23**	9	35
6	15	33	**24**	3	23
7	2	11	**25**	14	6
8	10	9	**26**	12	15
9	31	12	**27**	20	32
10	34	36	**28**	25	24
11	13	31	**29**	18	34
12	5	22	**30**	17	14
13	30	8	**31**	35	1
14	7	18	**32**	28	13
15	6	29	**33**	1	20
16	24	4	**34**	27	27
17	22	10	**35**	23	16
18	36	25	**36**	19	5

For the current study, the scales of the response Y values are different, so absolute relative error is calculated as (ARE)

$$\mathrm{ARE} = \frac{|\,\mathrm{Y} - \hat{\mathrm{Y}}\,|}{\mathrm{Y}} \tag{4.8}$$

to compare the performance of TreeMARS, TreeReg, CATreeMARS and CATreeReg without SUR and with SUR models, where $\hat{\mathrm{Y}}$ is the values of the predicted response. For example, $\hat{\mathrm{Y}}$ is $\hat{g}_{TreeMARS}(\mathbf{x})$ for TreeMARS. TreeReg and CATreeReg discussed above use the AIC algorithm from R. This dissertation also implements a stepwise regression method from the software SAS. The value of alpha-to-enter is 0.1, and the value of alpha-to-remove is 0.1.

For the models that use all the variables in the trees, the values of means, standard errors, and maximum of AREs are shown in Tables 4.37-4.40 (R: Response, TM: TreeMARS, TR: TreeReg, and RB: Both forward selection and backward elimination using R), and boxplots are shown in Figures 4.21-4.24. The means and standard deviations of ARE for TreeMARS with

SUR are smaller than TreeMARS without SUR. The maximum AREs for TreeMARS with SUR are also smaller than TreeMARS without SUR.

Using R and SAS, comparing the means of ARE of TreeReg with SUR and without SUR shows that the values are consistently smaller with SUR. Using R, the standard deviations of TreeReg with SUR for the first and second responses are smaller than TreeReg. Using SAS, the four standard deviations of TreeReg with SUR are smaller than TreeReg. TreeMARS has the maximum AREs except response 2. These boxplots shows all TreeMARS and TreeReg predictions are not significantly different. Many large outliers are shown for response 2 in Figure 4.22.

Table 4.37 Comparison of AREs for Six TM&TR Models for Response 1

	R1 TM	**R1 TM&SUR**	**R1 TR (RB)**	**R1 TR&SUR (RB)**	**R1 TR (SAS)**	**R1 TR&SUR (SAS)**
Mean	0.1177	0.1077	0.1326	0.1276	0.1307	0.1268
Standard Deviation	0.1160	0.1108	0.1141	0.1110	0.1064	0.1059
Maximum	0.4938	0.4869	0.4630	0.4619	0.4617	0.4571

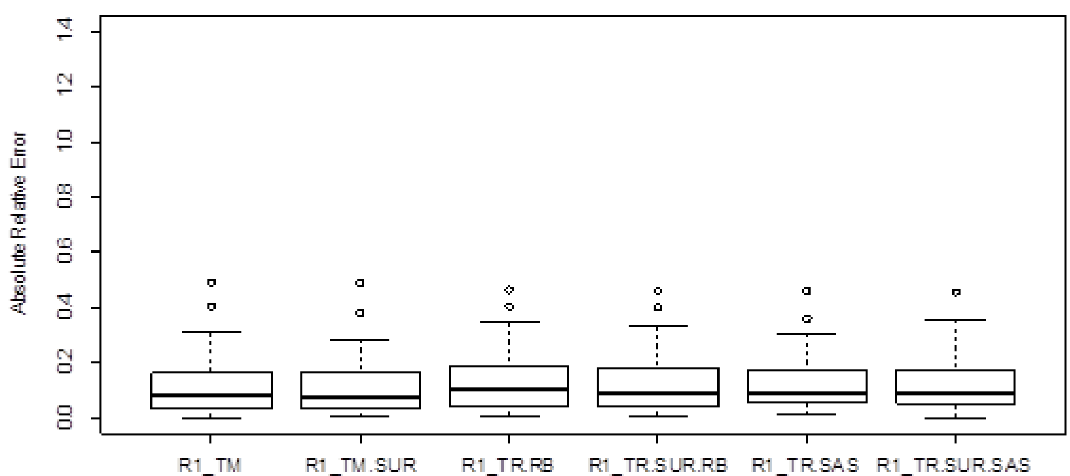


Figure 4.21 Comparison of Boxplots for Six TM&TR Models for Response 1

Table 4.38 Comparison of AREs for Six TM&TR Models for Response 2

	R2 TM	R2 TM&SUR	R2 TR (RB)	R2 TR&SUR (RB)	R2 TR (SAS)	R2 TR&SUR (SAS)
Mean	0.1734	0.1638	0.2597	0.2359	0.1992	0.1925
Standard Deviation	0.1662	0.1536	0.2547	0.2045	0.1828	0.1766
Maximum	0.9222	0.8550	1.1594	0.9250	0.8695	0.8692

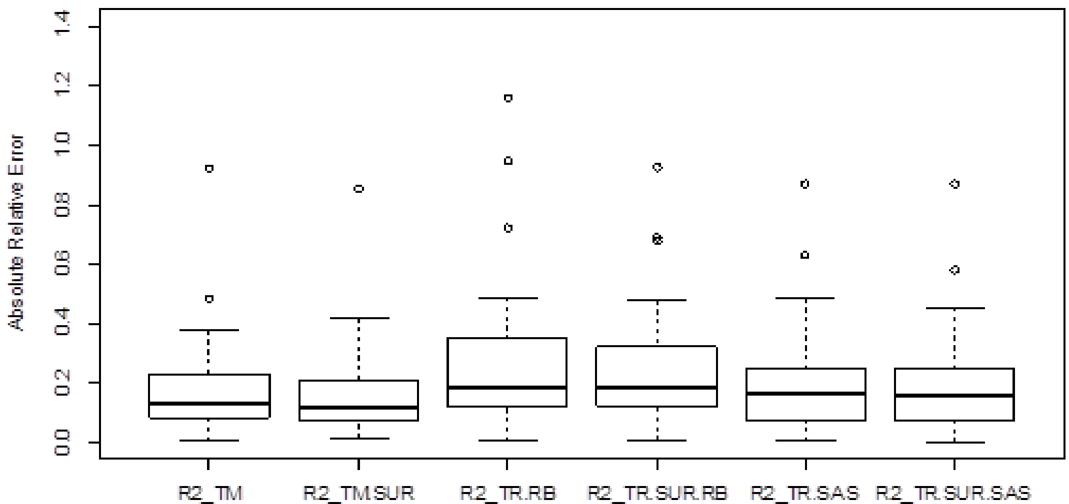


Figure 4.22 Comparison of Boxplots for Six TM&TR Models for Response 2

Table 4.39 Comparison of AREs for Six TM&TR Models for Response 3

	R3 TM	R3 TM&SUR	R3 TR (RB)	R3 TR&SUR (RB)	R3 TR (SAS)	R3 TR&SUR (SAS)
Mean	0.1139	0.1023	0.1244	0.1226	0.1257	0.1220
Standard Deviation	0.1120	0.1066	0.1009	0.1066	0.1023	0.1017
Maximum	0.4849	0.4631	0.3915	0.4373	0.4398	0.4353

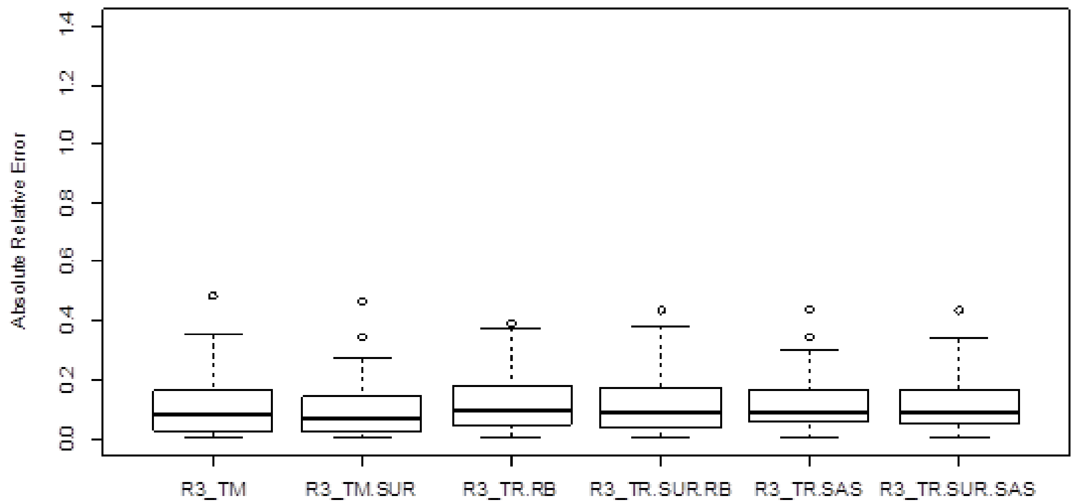


Figure 4.23 Comparison of Boxplots for Six TM&TR Models for Response 3

Table 4.40 Comparison of AREs for Six TM&TR Models for Response 4

	R4 TM	**R4 TM&SUR**	**R4 TR (RB)**	**R4 TR&SUR (RB)**	**R4 TR (SAS)**	**R4 TR&SUR (SAS)**
Mean	0.1138	0.1022	0.1243	0.1224	0.1256	0.1219
Standard Deviation	0.1119	0.1065	0.1008	0.1065	0.1022	0.1016
Maximum	0.4843	0.4625	0.3911	0.4368	0.4393	0.4347

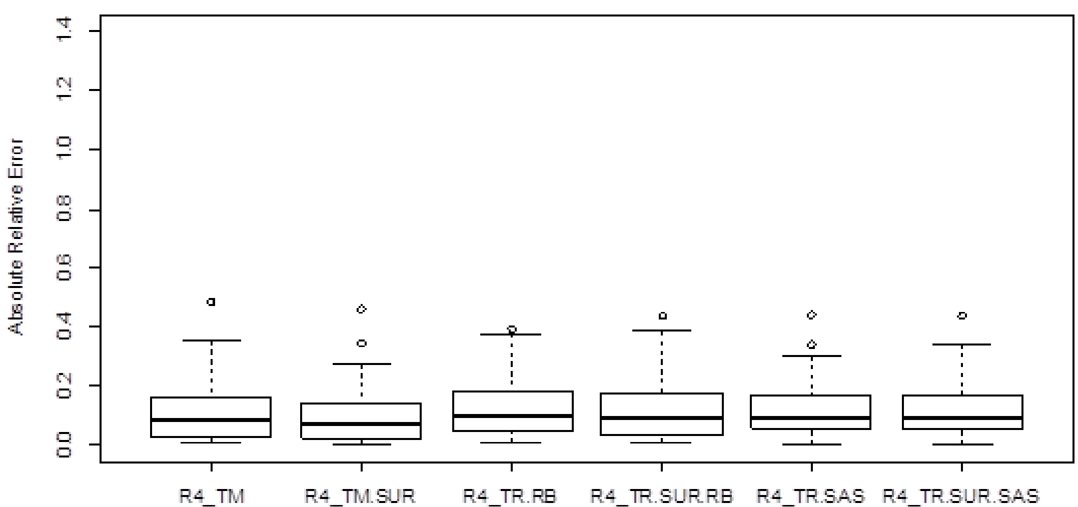


Figure 4.24 Comparison of Boxplots for Six TM&TR Models for Response 4

For the models that use all only the categorical variables in the trees, the values of means, standard errors, and maximum of AREs are shown in Tables 4.41-4.44 (CA: Categorical variables in tree model, TM: TreeMARS, and TR: TreeReg), and boxplots are shown in Figures 4.25-4.28. The means and standard errors of ARE for TreeMARS with SUR are smaller than TreeMARS without SUR. The maximum AREs occur for CATreeMARS. Moreover, CATreeReg with SUR using R has worse results for the third and fourth responses. CATreeReg with SUR using SAS has small mean and standard error values for all four responses. The boxplots show that all CATreeMARS and CATreeReg predictions are still not significantly different. Many large outliers are shown for response 2 in Figure 4.26.

Table 4.41 Comparison of AREs for Six CATM&CATR Models for Response 1

	R1 CATM	**R1 CATM &SUR**	**R1 CATR (RB)**	**R1 CATR &SUR (RB)**	**R1 CATR (SAS)**	**R1 CATR &SUR (SAS)**
Mean	0.2251	0.2181	0.1878	0.1580	0.1504	0.1423
Standard Deviation	0.1917	0.1785	0.1578	0.1197	0.1169	0.1124
Maximum	0.8700	0.7697	0.5595	0.4673	0.4397	0.4340

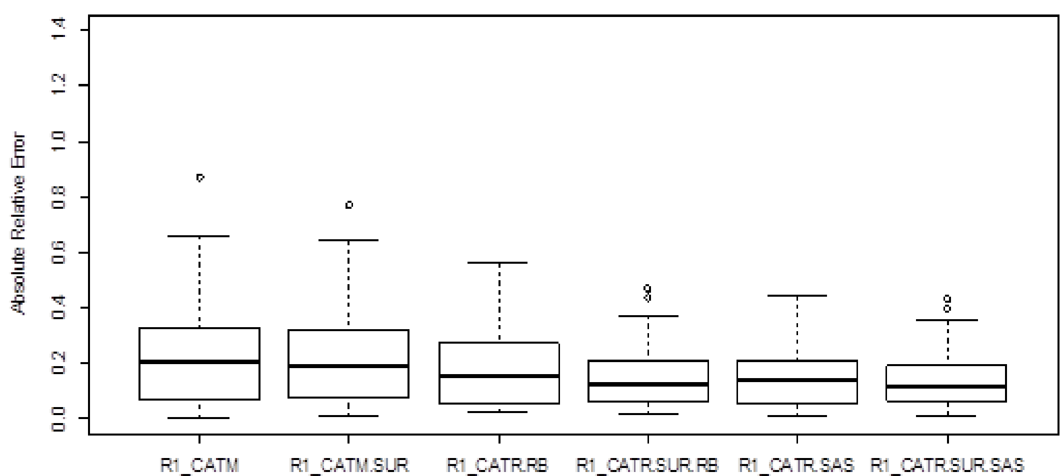


Figure 4.25 Comparison of Boxplots for Six CATM&CATR Models for Response 1

Table 4.42 Comparison of AREs for Six CATM&CATR Models for Response 2

	R2 CATM	R2 CATM &SUR	R2 CATR (RB)	R2 CATR &SUR (RB)	R2 CATR (SAS)	R2 CATR &SUR (SAS)
Mean	0.2814	0.2694	0.3071	0.3014	0.2518	0.2335
Standard Deviation	0.3239	0.3047	0.2473	0.2507	0.2151	0.1976
Maximum	1.3667	1.2828	1.0039	1.0376	1.0526	0.9217

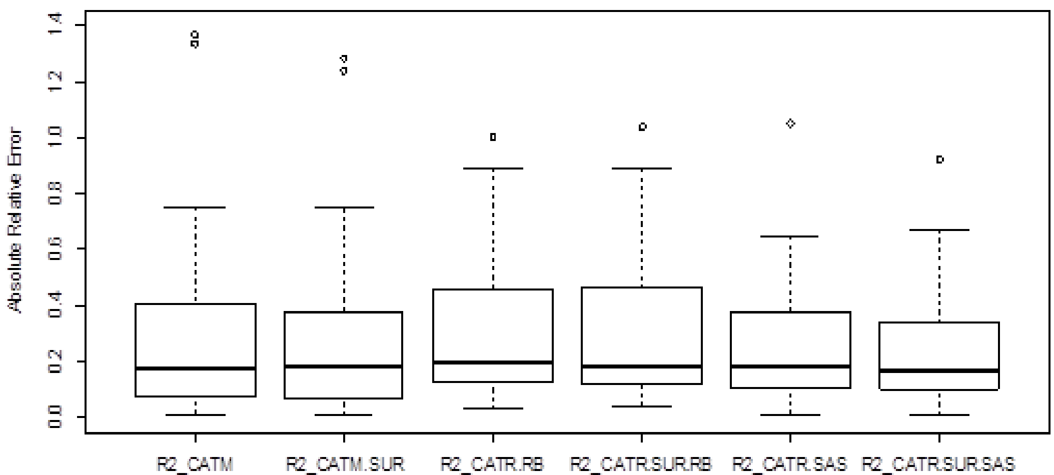


Figure 4.26 Comparison of Boxplots for Six CATM&CATR Models for Response 2

Table 4.43 Comparison of AREs for Six CATM&CATR Models for Response 3

	R3 CATM	R3 CATM &SUR	R3 CATR (RB)	R3 CATR &SUR (RB)	R3 CATR (SAS)	R3 CATR &SUR (SAS)
Mean	0.2025	0.1859	0.1671	0.1723	0.1594	0.1412
Standard Deviation	0.1700	0.1583	0.1362	0.1395	0.1285	0.1124
Maximum	0.6269	0.6090	0.5207	0.5803	0.5535	0.4176

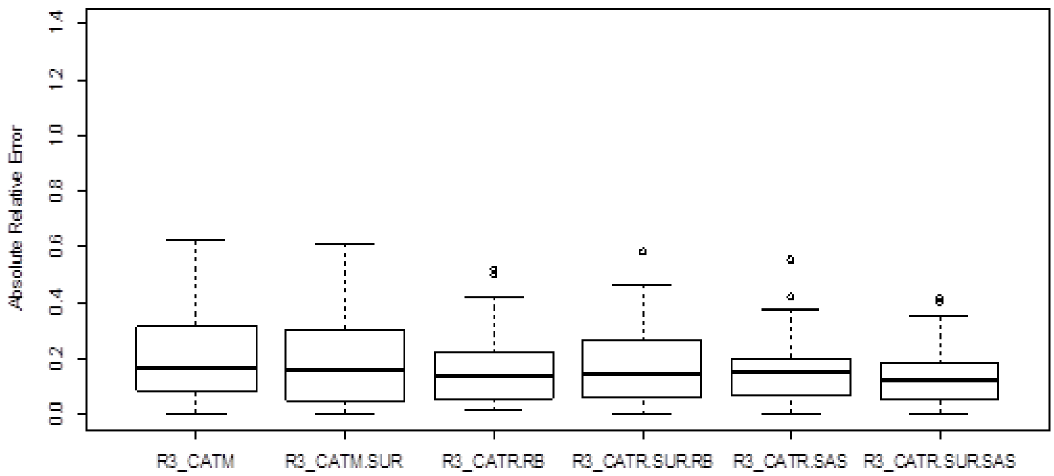


Figure 4.27 Comparison of Boxplots for Six CATM&CATR Models for Response 3

Table 4.44 Comparison of AREs for Six CATM&CATR Models for Response 4

	R4 CATM	**R4 CATM &SUR**	**R4 CATR (RB)**	**R4 CATR &SUR (RB)**	**R4 CATR (SAS)**	**R4 CATR &SUR (SAS)**
Mean	0.2022	0.1857	0.1659	0.1721	0.1593	0.1411
Standard Deviation	0.1699	0.1582	0.1365	0.1394	0.1284	0.1122
Maximum	0.6262	0.6083	0.5194	0.5799	0.5529	0.4171

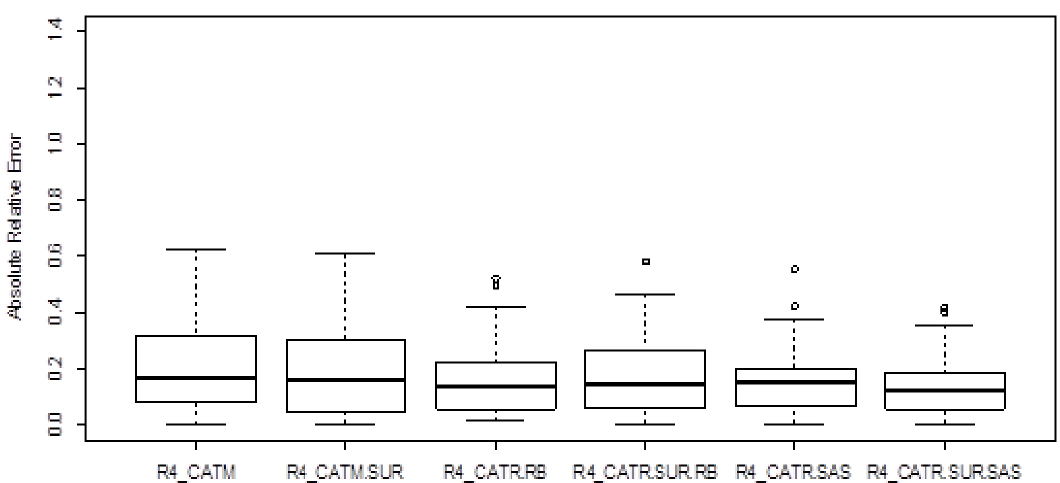


Figure 4.28 Comparison of Boxplots for Six CATM&CATR Models for Response 4

CHAPTER 5

CONCLUSIONS AND FUTURE WORK

5.1 Conclusions

Green building enhances the quality and protection of the environment in which people work and live, and the MSMO decision-making framework is proposed to enable more comprehensive guidance for builders. This dissertation employed the software eQUEST, which provides enough building options and materials to evaluate building performance from an energy standpoint, and design of experiments is used to organize a set of eQUEST runs. In the first study, the eQUEST output using a two-level experimental design with 216 runs was analyzed using various multivariate (multi-response) methods, and most factor variables were identified as important.

This dissertation describes the green building case study and the framework, developed in collaboration with the green building expert, Anthony Robinson. The discussions included activity percent area specifications, first building cost and annual maintenance cost in the residential low-rise building using eQUEST simulation, and the order of twelve stages that was organized for decision-making framework.

In the second study, an experimental design method was developed for a mix of discrete-categorical, discrete-numerical, and continuous factor variables. This second design was combined with the first design, and both TreeReg and TreeMARS models were constructed. In Sahu (2011), TreeMARS was shown to have better prediction accuracy than TreeReg with single response. Because green building research involves multiple performance metrics, this dissertation extended TreeReg and TreeMARS to the multiple response case by using SUR. Since green building performance metrics are likely correlated, SUR is more appropriate than OLS that processes the responses separately. In the green building case study, SUR demonstrates consistently lower standard errors, but the predictor errors for TreeMARS, TreeReg, CATreeMARS and CATreeReg are not significantly different, but TreeMARS and TreeReg perform better than CATreeMARS and CATreeReg. Overall, TreeMARS with SUR shows the smallest mean for AREs.

5.2 Future Work

In future work, other multi-response methodologies, such as C&W and MD, will be compared. Additional analyses using statistical data mining methods [81] and larger experimental designs should be conducted, including further study of the combined design approach in Section 4.1. The resulting models need to be analyzed for practical relevance, as attempted in Section 4.4. To supplement eQUEST, computer runs of ATHENA will be used to quantify environmental impact. Finally, there are other building types, including commercial low rise, commercial multi-story, industrial low rise, and office building. To accommodate more general building structures, a modified version of Table 3.1 is shown in Table 5.1, based the suggestions by green building expert, Anthony Robinson.

Table 5.1 Revised Stages and Decision Variables for Green Building

Stage	Building Stage with Options
1	**Siting Options** ● Orientation and Footprint (eQUEST)
2	**Foundation System** ● Concrete Ground Floor (eQUEST) ● Concrete Slab on Grade (ATHENA) ● Generic Portland Cement (BEES) ● Steel Foundation System
3	**Wells and Septic System** ● Concrete Septic Tank ● Fiberglass Septic Tank
4	**Wall System** ● Concrete Wall (ATHENA, BEES, eQUEST) ● Curtain Wall (ATHENA) ● Drywall ● Metal Frame (eQUEST) ● Straw Bale Walls ● Wood Frame (eQUEST)
5	**Roof System** ● Concrete Tile Roof (ATHENA, eQUEST) ● Generic Fiber Cement Roof (BEES)

Table 5.1 – *Continued*

	● Roof Surface Materials (eQUEST)
6	**Window System** ● Clear/Tint Windows (eQUEST) ● Glazed Windows ● Low-e Windows (eQUEST) ● Reflective Windows (eQUEST) ● Wood Frame Windows (ATHENA, eQUEST)
7	**Door System** ● Steel Door (ATHENA, eQUEST) ● Wood Door (eQUEST)
8	**Plumbing System** ● Freshwater System ● Greywater System ● Rainwater Catchment System
9	**Electrical System** ● AC System (eQUEST) ● Both AC and Solar System ● Solar System
10	**Ventilation System** ● Balanced Ventilation System ● Exhaust Ventilation System ● Supply Ventilation System ● Ventilation-Activity Areas (eQUEST)
11	**Heating and Cooling System** ● Fan System (eQUEST) ● HVAC System (eQUEST)
12	**Landscaping System** ● Sprinkler System

REFERENCES

[1] Zhang, Z., Wu, X., Yang, X. and Zhu, Y. (2006). "BEPAS—a life cycle building environmental performance assessment model," *Building and Environment*, 41, pp. 669–675.

[2] Retzlaff, R. C. (2008). "Green building assessment systems: a framework and comparison for planners," *Journal of the American Planning Association*, 74, pp. 505–519.

[3] U.S. Environmental Protection Agency (USEPA). http://www.epa.gov/ (accessed February, 2012).

[4] U.S. Department of Energy (USDOE). http://energy.gov/ (accessed February, 2012).

[5] American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE). http://www.ashrae.org/ (accessed February, 2012).

[6] Leadership in Energy and Environmental Design (LEED) (2012). U.S. Green Building Council (USGBC). http://www.usgbc.org/DisplayPage.aspx?CMSPageID=222 (accessed February, 2012).

[7] Energy Star (2011). U.S. Environmental Protection Agency and the U.S. Department of Energy, USA. http://www.energystar.gov/index.cfm?c=home.index (accessed December, 2011).

[8] NAHBGreen (2011). National Association of Home Builders, USA. http://www.nahbgreen.org/ (accessed December, 2011).

[9] Sacks, J., Welch, W. J., Mitchell, T. J. and Wynn, H. P. (1989). "Design and analysis of computer experiments," *Statistical Science*, 4, pp. 409–423.

[10] Building Energy Software Tools Directory (2011). Energy Efficiency and Renewable Energy (EERE), USA. http://www.eere.energy.gov/ (accessed December, 2011).

[11] ATHENA Impact Estimator for buildings, V. 4.1.13 (2011). The ATHENA Institute, Canada.

[12] BEES Online (2010). National Institute of Standards and Technology, USA.

[13] eQUEST, V. 3.63b (2009). James J. Hirsch and Associates, USA.

[14] Wang, W., Zmeureanua, R. and Rivard, H. (2005). “Applying multi-objective genetic algorithms in green building design optimization,” *Building and Environment*, 40, pp. 1512–1525.

[15] Castro-Lacouture, D., Sefair, J. A., Flórez, L. and Medaglia, A. L. (2009). “Optimization model for the selection of materials using a LEED-based green building rating system in Colombia,” *Building and Environment*, 44, pp. 1162–1170.

[16] Nielsen, T. R. (2002). “Optimization of buildings with respect to energy and indoor environment,” Ph.D. Dissertation, Technical University of Denmark.

[17] Osman, A., Norman, B. A. and Ries, R. (2008). “Life cycle optimization of building energy systems,” *Engineering Optimization*, 40, pp. 157–178.

[18] Hasan, A., Vuolle, M. and Sirén, K. (2008). “Minimisation of life cycle cost of a detached house using combined simulation and optimization,” *Building and Environment*, 43, pp. 2022–2034.

[19] Building Energy Optimization (BEopt). http://beopt.nrel.gov/ (accessed February, 2012).

[20] Center for sustainable systems (2010). University of Michigan. http://css.snre.umich.edu/ (accessed December, 2010).

[21] Johnson, R. A. and Wichern, D. W. (2007). *Applied multivariate statistical analysis*. Pearson Prentice Hall, New Jersey.

[22] Derksen, S. and Keselman, H. J. (1992). “Backward, forward and stepwise automated subset selection algorithms: frequency of obtaining authentic and noise variables,” *British Journal of Mathematical and Statistical Psychology*, 45, pp. 265–282.

[23] Copas, J. B. (1983). “Regression, prediction and shrinkage,” *Journal of the Royal Statistical Society B*, 45, pp. 311–354.

[24] Tibshirani, R. (1996). “Regression shrinkage and selection via the lasso,” *Journal of the*

Royal Statistical Society B, 58, pp. 267–288.

[25] Zellner, A. (1962). “An efficient method of estimating seemingly unrelated regression equations and test for aggregation bias,” *Journal of the American Statistical Association*, 57, pp. 348–368.

[26] Henningsen, A. and Hamann, J. D. (2007). “Systemfit: a package for estimating systems of simultaneous equations in R” *Journal of Statistical Software*, 23, pp. 1–40.

[27] Shah, H. K., Montgomery, D. C. and Carlyle, W. M. (2004). “Response surface modeling and optimization in multiresponse experiments using seemingly unrelated regressions,” *Quality Engineering*, 16, pp. 387–397.

[28] Breiman, L. and Friedman, J. H. (1997). “Predicting multivariate responses in multiple linear regression,” *Journal of Royal Statistical Society B*, 59, pp. 3–54.

[29] Srivastava, Muni S. and Solanky, T. K. S. (2003). “Predicting multivariate response in linear regression model,” *Simulation and Computation*, 32, pp. 389–409.

[30] Xu, Q. S., de Jong, S., Lewi, P. and Massart, D. L. (2004). “Partial least squares regression with Curds and Whey,” *Chemometrics and Intelligent Laboratory Systems*, 71, pp. 21–31.

[31] Breiman, L., Friedman, J. H., Olshen, R. A. and Stone, C. J. (1984). *Classification and Regression Trees*, Chapman & Hall/CRC.

[32] R. http://www.r-project.org/ (accessed December, 2011).

[33] Package tree. http://cran.r-project.org/web/packages/tree/tree.pdf (accessed December, 2011).

[34] Package rpart. http://cran.r-project.org/web/packages/rpart/rpart.pdf (accessed December, 2011).

[35] Friedman, J. H. (1991). “Multivariate adaptive regression splines,” *Annals of Statistics*, 19, pp. 1–141.

[36] Loh, W.-Y. (2002). “Regression trees with unbiased variable selection and interaction detection,” *Statistica Sinica*, 12, pp. 361–386.

[37] Kim, H., Loh, W.-Y., Shih, Y.-S. and Chaudhuri, P. (2007). "Visualizable and interpretable regression models with good prediction power," *IIE Transactions*, 39, pp. 565–579.

[38] De'ath, G. (2002). "Multivariate regression trees: a new technique for constrained classification analysis," *Ecology*, 83, pp. 1103–1117.

[39] Multivariate partitioning (MVPART). http://cran.r-project.org/web/packages/ mvpart/index.html (accessed December, 2011)

[40] Loh, W.-Y. (2011). GUIDE V. 11.3, Classification and Regression Trees.

[41] Multivariate adaptive regression splines. http://en.wikipedia.org/ wiki/ Multivariate_adaptive_regression_splines (accessed December, 2011)

[42] Sekulic, S. and Kowalski, B. R. (1992). "MARS: a tutorial," *Journal of Chemom*etrics, 6, pp. 199–216.

[43] Hastie, T., Tibshirani, R. and Buja, A. (1994). "Flexible discriminant analysis by optimal scoring," *JASA*, pp. 1255–1270.

[44] Linear discriminant analysis from Wikipedia. http://en.wikipedia.org/wiki/ Linear_discriminant_analysis (accessed December, 2011)

[45] Friedman, J. H. and Tukey, J. W. (1974). "A projection pursuit algorithm for exploratory data analysis," *IEEE Transactions on Computers*, C-23, pp. 881–890.

[46] Friedman, J. H. (1987). "Exploratory projection pursuit," *Journal of the American Statistical Association*, 82, pp. 249–266.

[47] Friedman, J. H. and Stuetzle, W. (1981). "Projection pursuit regression," *Journal of the American Statistical Association*, 76, pp. 817–823.

[48] Friedman, J. H. (1985). "Classification and multiple regression through projection pursuit," Technical Report, LCS 12, pp. 1–31.

[49] Friedman, J. H. (1984). "SMART users guide," Technical Report, LCS 1, pp. 1–24.

[50] Frank, I. E. (1995). "Tutorial modern nonlinear regression methods," *Chemometrics and Intelligent Laboratory Systems*, 27, pp. 1–9.

[51] Alexander, W. P. and Grimshaw, S. D. (1996). “Treed regression," *The Journal of Computational and Graphical Statistics*, 5, pp. 156–175.

[52] Sahu, S. (2011). “Multivariate adaptive regression spline based framework for statistically parsimonious adaptive dynamic programming,” Ph.D. Dissertation, The University of Texas at Arlington.

[53] Pan, J.-N., Pan J. and Lee, C.-Y. (2009). “Finding and optimizing the key factors for the multiple-response manufacturing process,” *International Journal of Production Research*, 47, pp. 2327–2344.

[54] Taguchi, G. and Jugulum, R. (2002). “The Mahalanobis-Taguchi strategy: a pattern technology system,” New York: John Wiley & Sons.

[55] Manne, R. (1987). Analysis of two partial-least-squares algorithms for multivariate calibration, *Chemometrics and Intelligent Laboratory Systems*, 2, pp. 283–290.

[56] Wold, S., Wold, N. K. and Skagerberg, B. (1989). “Nonlinear PLS modeling,” *Chemometrics and Intelligent Laboratory Systems*, 7, pp. 53–65.

[57] Baffi, G., Martin, E. B. and Morris, A. J. (1999). “Non-linear projection to latent structures revisited: the quadratic PLS algorithm,” *Computers in Chemical Engineering*, 23, pp. 95–411.

[58] Qin, S. J. and McAvoy, T. J. (1992). “Nonlinear PLS modeling using neural networks,” *Computers and Chemical Engineering*, 16, pp. 379–391.

[59] Baffi, G., Martin, E. B. and Morris, A. J. (1999). “Non-linear projection to latent structures revisited: the neural network PLS algorithm,” *Computers and Chemical Engineering*, 23, pp. 1293–1307.

[60] Wold, S., Sjöström, M. and Eriksson, L. (2001). “PLS-Regression: a basic tool of Chemometrics,” *Chemometrics and Intelligent Laboratory Systems*, 58, pp. 109–130.

[61] Steps to Building a House. http://www.byoh.com/stepbystep.htm (accessed December, 2011)

[62] Birge, J. R. and Louveaux, F. (1997). *Introduction to stochastic programming*. New York: Springer-Verlag.

[63] Chen, V. C. P. (2001). “Measuring the goodness of orthogonal array discretizations for stochastic programming and stochastic dynamic programming,” *SIAM Journal of Optimization*, 12, pp. 322–344.

[64] Tsai, J. C. C., Chen, V. C. P., Beck, M. B. and Chen, J. (2004). “Stochastic dynamic programming formulation for a wastewater treatment decision-making framework,” *Annals of Operations Research*, 132, pp. 207–221.

[65] Hedayat, A. S., Sloane, N. J. A. and Stufken, J. (1999). *Orthogonal arrays: theory and applications*. New York: Springer-Verlag.

[66] R, V. 2.11.1 (2010). DoE.base: full factorials, orthogonal arrays and base utilities for DoE packages.

[67] Crawl Space Foundation. http://www.byoh.com/crawlspaces.htm (accessed February, 2012).

[68] Crawl Space Foundation vs. Slab Foundation. http://www.byoh.com/crawl-space-foundation-versus-slab-foundation.htm (accessed February, 2012).

[69] SAS/STAT® 9.2 User's Guide (2008). SAS Institute Inc., Cary, NC, USA.

[70] Kung, P., Chen, V. C. P. and Robinson, A. (2011). “Multivariate modeling for a multi-stage green building framework,” *IEEE International Symposium on Sustainable Systems and Technology*, pp. 1–6.

[71] Kung, P., Chen, V. C. P. and Robinson, A. (2011). “A multivariate analysis of green building options,” *Clean Technology and Sustainable Industries Organization*, Chapter 8, pp. 363–366.

[72] Salford Systems. http://www.salford-systems.com/ (accessed February, 2012).

[73] Insulation R-Value. http://www.nachi.org/insulation-r-value.htm (accessed February, 2012).

[74] R-Value Table. http://www.coloradoenergy.org/procorner/stuff/r-values.htm (accessed February, 2012).

[75] R-values of Insulation and Other Building Materials. http://archtoolbox.com/materials-systems/thermal-moisture-protection/24-rvalues.html (accessed February, 2012).

[76] R-value (insulation). http://en.wikipedia.org/wiki/R-value_%28insulation%29 (accessed February, 2012).

[77] SOBOL_DATASET Generate Sobol Datasets. http://people.sc.fsu.edu/~jburkardt/m_src/sobol_dataset/sobol_dataset.html (accessed May, 2011).

[78] Stepwise Regression. http://en.wikipedia.org/wiki/Stepwise_regression (accessed February, 2012).

[79] R: Choose a model by AIC in a Stepwise Algorithm. http://stat.ethz.ch/R-manual/R-devel/library/stats/html/step.html (accessed February, 2012).

[80] Meckesheimer, M., Booker, A. J., Barton R. R. and Simpson, T. W. (2002). "Computationally inexpensive metamodel assessment strategies," *AIAA Journal*, 40, pp. 2053–2060.

[81] Hastie, T., Tibshirani, R. and Friedman, J. H. (2001). *The elements of statistical learning: data mining, inference, and prediction*. New York: Springer-Verlag.

BIOGRAPHICAL INFORMATION

Pin Kung received a Bachelor’s degree in Statistics and minored in Industrial Management Science from National Cheng Kung University in Taiwan. He later received a Master’s degree in Industrial Engineering and Management from Yuan Ze University in Taiwan. In 2007, he enrolled in the Ph.D. program in Industrial and Manufacturing Systems Engineering at The University of Texas at Arlington. During his doctoral studies, he was a graduate teaching assistant at UTA. His research interests include Statistics, Green Building, Quality Control, and Fuzzy Control.